2011

Experience of Supervision Scale: The Development of an Instrument to Measure Child Welfare Workers' Experience of Supervisory Behaviors

Mary Parente
Virginia Commonwealth University

Experience of Supervision Scale: The Development of an Instrument to Measure Child Welfare Workers' Experience of Supervisory Behaviors

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

by

Mary E. Parente
BA, Cornell University, 1989
MSW, Boston College, 1993

Director: Patrick V. Dattalo, PhD
Associate Professor
School of Social Work

Virginia Commonwealth University
Richmond, VA
May, 2011

Acknowledgment

I would first like to thank my dissertation chair, Pat Dattalo, who remained a calm and steady source of support throughout the eight long years it has taken me to finish this project. Each time I came to him with another question or in need of advice, he was unfailingly encouraging and helpful. He was also endlessly patient with my second guessing my statistical knowledge and repeated misspelling of statistical terminology. He read everything I sent him, and gave exactly the feedback I needed to be able to move to the next stage of my project.

I would also like to express appreciation to my committee members: Mary Secret, Jenny Jones, and Janet Hutchinson. Each made a considerable commitment to me and stuck by me throughout. I am grateful to each for their willingness to meet with me whenever I needed their input and for their contributions to the dissertation process.

I could never have finished without the constant support and encouragement I received from my supervisors and coworkers at Charlottesville Department of Social Services. Even when I wasn't sure I could do it, they never doubted me, and when I needed to take an extended "vacation" to focus on writing, they covered for me with enthusiasm.

I want to thank my friends and colleagues from the PhD program, especially Gerry Meeks and Allison Sampson, who understood all too well the pressures of balancing academic pursuits, work, and parenting.

Finally, I want to thank my family for their practical assistance as well as their unwavering support. My children, Nikki and Tim, have done their best to tolerate my crazy schedule and periods of "crankiness." Over the years it has taken me to finish this, they have grown into enjoyable and admirable young adults. My parents provided me with a healthy, happy childhood home and the many opportunities which contributed to my becoming the person I am today. I was so lucky to get you as parents! And to my sister, Peggy, proofreader extraordinaire, without your matter-of-fact willingness to read whatever I sent you whenever I needed you to, I really might not have gotten it done. Knowing that you were waiting to help weed out the typos gave me the courage to get the words on the page to begin with.

Table of Contents

List of Tables

Abstract

EXPERIENCE OF SUPERVISION SCALE: THE DEVELOPMENT OF AN INSTRUMENT TO MEASURE CHILD WELFARE WORKERS' EXPERIENCE OF SUPERVISORY BEHAVIORS

By Mary E. Parente, PhD

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2011

Director: Patrick V. Dattalo, Associate Professor, School of Social Work

The purpose of this study was to develop and test the Experience of Supervision Scale, which measures the experience of supervisory behaviors from the perspective of child welfare workers. Child welfare supervision is gaining increased attention as states struggle to improve performance measures of their child welfare systems. However, the lack of a standardized instrument which measures supervision through an analysis of supervisory behaviors from the supervisee's perspective has hindered efforts in these areas. The Experience of Supervision Scale is based on Kadushin's three function casework supervision model of supervision (administrative, educational, and supportive) which was chosen as the theoretical framework due to this model's emphasis on public agency clinical case work practice. Scale items were generated through focus groups of child welfare workers and supervisors. Items were reviewed by an expert panel and the scale was then administered to a statewide sample of 165 child welfare workers. The refined Experience of Supervision Scale consists of 32 items. In this

study, it demonstrated high internal consistency with a Cronbach's alpha of .962. Factor analysis yielded a three factor solution consistent with the casework model of supervision's structure, while also suggesting slight modifications to the model. The three factors derived included: Performance Support, Emotional support/Availability, and Administrative activities. The results of this study suggest that the Experience of Supervision Scale has appropriate psychometric properties for use in complex research in which casework supervision is a variable.

Chapter 1: Introduction

The supervision of child welfare staff is increasingly being recognized to be of critical importance in insuring the quality of case work services that vulnerable children and their families receive through the child welfare system (National Child Welfare Resource Center for Organizational Improvement, 2007). As a field, child welfare is plagued by cumbersome bureaucracies, lack of resources, complex and high risk situations, and difficulties related to maintaining a well-prepared and highly trained workforce of child welfare social workers adequate to meet the demands of the system (Balfour & Neff, 1993; Tracy & Pine, 2000; United States General Accounting Office [US GAO], 2003; US GAO, 2007). Supervision has been found to be of critical importance in the retention of child welfare workers (Cyphers, 2003; Samantrai, 1992; US GAO, 2003, 2007) and in the development in those workers of the necessary knowledge, skills, and abilities to do the job well (Bednar, 2003; Cearley, 2004; Wehrman, Shin, & Poertner, 2002). Using child welfare supervision as an intervention into the child welfare system is gaining increased attention as states struggle to improve performance measures of their child welfare systems (Bednar, 2003; Cearley, 2004; Collins-Camargo, 2006; Faller, 2004; Griffin & Shiell, 2003; Jacquet, Clark, Morazes, & Withers, 2007; National Child Welfare Resource Center for Organizational Improvement, 2007; Renner, Porter, & Priester, 2009; Smith, 2005; Strand & Badger, 2005, 2007; Travis & Mor Barak, 2010; Zlotnik, DePanfilis, Daining, & Lane, 2005). However, the process of supervision in child welfare has not been studied extensively.

What is it that child welfare supervisors are doing well or not well? How do those things influence the practice of the workers they supervise? What specific supervisory behaviors are experienced as helpful? What does a supervisor do or fail to do that results in a worker experiencing supervision as unhelpful? Or unsupportive? What specific supervisory behavioral changes make the biggest difference in workers' experiences of supervision? Or contribute to changes in workers' practice? Or to reducing preventable turnover?

There are very few well-developed empirical studies which focus specifically on child welfare supervision, and studies of child welfare workforce issues which include supervision as a variable, operationalize supervision inconsistently (Zlotnik et al., 2005). As a result, little is known about the practice of supervision in the child welfare agency, especially from the perspective of the child welfare worker. In order to improve the quality of child welfare supervision and ultimately the quality of services provided to vulnerable children and families being served by the system, it is necessary to better understand the process of supervision, especially, supervisory behaviors and how workers experience them. A standardized, valid instrument that adequately measures the behaviors of child welfare supervisors in the provision of supervision from the perspective of child welfare workers would be a valuable tool in this endeavor.

Child Welfare Supervision as an Intervention

Much of the federal funding and federal policy guidelines for child welfare services administered by the states are established through the requirements in Title IV-B and IV-E of the Social Security Act (SSA). Amendments to the SSA in 1994 authorized the United States Department of Health and Human Services (HHS) to review the states' child and family service programs for conformity with federal policy requirements. Typically these reviews consisted of

federal audits of states' written procedural policies. However, in accordance with the provisions set out in the Adoption and Safe Families Act (ASFA) of 1997, in 2000 the HHS established new procedures for reviewing the states' child welfare programs. The Child and Family Services Reviews (CFSRs), conducted by the Children's Bureau are intended 1) to monitor the states' level of compliance with federal policies, 2) to assess how children and families actually fare in the states' child welfare programs, and 3) to help the states to improve their capacity to achieve positive outcomes for children and families. The CFSRs include state level data concerning practice outcomes, case record review, and interviews with children and families receiving services as well as with foster families, child welfare workers, and service providers. All 50 states and Puerto Rico completed their first CFSR by 2004. Each was evaluated in terms of seven outcome areas and seven systemic factors. Not one state was found "to be in substantial conformity" with all the performance goals evaluated. Each state has since developed and implemented a program improvement plan (PIP) focused on improving the state's performance in the next round of reviews. This second round of CFSRs began in 2007 (HHS, n.d. a).

Review of the findings from the initial round of the CFSRs yielded common challenges among the states. One of the most significant findings was the degree to which staffing shortages and high caseloads impact the states' abilities to address those challenges (HHS, n.d. b; US GOA, 2003). In 2002, US GAO researchers analyzed the 27 CFSRs which had been completed by the Department of Health and Human Services. No state reviewed had met all of the performance goals established by the federal government through the Adoption and Safe Families Act. The researchers found that in all 27 CFSRs completed at that time, workforce deficiencies including high caseloads, training deficiencies, and staffing shortages, contributed to failure to attain at least one performance measure. Two states, Georgia and Oregon, showed

workforce deficiencies contributing to the failure to attain 14 and 16 measures, respectively (US GAO, 2003).

For the last 35 states reviewed in the first round of CFSRs, issues around new worker training and the provision of ongoing staff training were found to be problematic in nearly 50% of the states. Individual case reviews specifically identified problems with workers' failure to include families, especially fathers, and older children in case planning; to adequately assess mental needs and provide services to children; and to adequately assess family issues which may contribute to child maltreatment including substance abuse, mental illness, and domestic violence in as many as 63% of the states. More frequent, face-to-face contact between child welfare workers and children was found to be significantly associated with providing adequate services to prevent removal, establishing permanency goals for children in foster care in a timely manner, and achieving permanency for children in foster care in a timely manner (HHS, n.d. c).

Federal guidance has resulted in states instituting policies which specifically address particular areas of federal concerns arising from the CFSRs. For example, the Child and Family Services Act of 2006 requires states to implement mandated minimum face-to-face contact between foster care caseworkers and children in foster care. Policy implementation guidelines from the HHS's Administration for Children and Families specify a timeline for implementation, requirements around states' reporting data regarding monthly caseworker contact, and penalties both for failing to report and for failing to meet increasingly challenging benchmarks towards the final goal of full compliance by October 1, 2011. At that time, all states must show that face-to-face caseworker contacts occurred for 90% or better of all cases and that more than 50% of these contacts occurred in the child's place of residence (HHS, 2008).

In Virginia, a foster care policy Broadcast was released in August, 2008, spelling out the state's plans to work towards compliance with this federal mandate (Virginia Department of Social Services [VDSS], July, 2008). It is important to note, however, that very limited federal funding is being provided to assist the states to meet the October 1, 2011, deadline. According to the Administration for Children and Families ACYF-CB-IM-07-05 program instructions dated February 28, 2007, in 2006, $40 million was made available to the states with the requirement that their share "must be used to support monthly caseworker visits with children who are in foster care under the responsibility of the State with a primary emphasis on activities designed to improve caseworker retention, recruitment, training and ability to access the benefits of technology." Access to the funds requires a 25% state match (HHS, 2007). In FY 2010, the Monthly Caseworker Visit federal allotment provided to the states totaled a mere $20 million (HHS, 2009).

Clearly, lower caseloads, additional staff and additional resources would be helpful in improving the frequency of contact between foster care workers and children in foster care (US GAO, 2003, 2007). However, the funds made available to the staff are not sufficient or long-standing enough to enable the addition of positions. Virginia plans to use its money to reimburse local Department of Social Services for the costs associated with foster care workers traveling to see children in foster care who are placed out of state or at substantial distances from their communities within the state. The sum available was not considered sufficient even to address the need for improved, portable technology to insure the documentation of worker visits (VDSS, August, 2008).

The penalties for failing to substantially conform to CFSR performance measures are still not entirely clear. The Adoption and Safe Families Act which lead to the institution of the

CFSRs, also modified the SSA to include provisions for the development of a "performance-based incentive system" whereby states can be penalized through the withholding of Title IV-E and IV-B funds for failing to meet performance standards (US GAO, 2006). As specific requirements for the states are initiated, specific penalties are better defined. For example, the Administration for Children and Families April 18, 2008, program instruction ACYF-CB-PI-08-03 includes the following penalties for states falling short of the target percentage for monthly face-to-face contacts: 1) less than 10% short will result in a 1% decrease in the federal Title IV-E and IV-B match rate (currently 75% to the state's 25%); 10 to 19% short will result in a 3% decrease; and 20% or more short will result in a 5% decrease. These penalties will be instituted in each year between 2009 and 2011 as the states' progress towards meeting the 90% goal is assessed. A technical bulletin issued in 2010 additionally notes that should states fail to correct the problems addressed in their first CFSR review, when the second CFSR review is conducted the penalties will be doubled for each of the 7 outcomes and 7 systemic factors that continues in nonconformity (HHS, 2010 c).

Increasingly then, individual states are left to wrestle with the question: how can we do more with limited additional resources? The ultimate goal of an improved child welfare workforce through decreased turnover and improved educational and skill levels appears unlikely to be facilitated through additional federal funding. But failure to improve the quality of services provided to children and families and to meet federal performance measures is likely to result in substantial financial penalties.

How then can states influence the quality of their workforce in order to improve the quality of the services provided to at-risk children and families? In 2007, the National Child Welfare Resource for Organizational Improvement (NRCOI) released a CFSR training and

technical assistance (T/TA) package. The material is intended to help agencies prepare for the second round of CFSRs. Focus Area Three of the T/TA is "Strengthening Child Welfare Supervision as a Key Practice Change Strategy." Supervisors are identified as key to changing ongoing practice in the agency and thereby improving outcomes for children and families. Three factors are seen as contributing to the importance of the child welfare supervisor as a "practice change agent": 1) hierarchically, they are positioned between administration and frontline workers and therefore have access to administrative data which can be used to guide practice; 2) they provide a highly visible role model for workers; and 3) they are able to provide clinical direction to workers which will guide the workers' practice. The use of administrative data to guide practice and the provision of clinical supervision, particularly, are seen as a "re-design" and "re-conceptualization" of child welfare supervision. The following list of activities is the suggested strategy for using data and [clinical] supervision to change practice:

1) Target individual cases: identify specific cases on which to take action (clinical supervision)
2) Modify agency systems: seek to change or influence changes in agency administrative systems
3) Create key actor collaborations: seek input, create collaborations, or influence systems that impact the agency's ability to achieve outcomes
4) Develop staff capacity: build skills and knowledge that enable staff to achieve outcomes regardless of function
5) Secure resources: obtain appropriate resources (e.g. staff, financial) to better achieve outcomes

6) Reward others: reward and motivate staff and key actors when they achieve optimal outcome levels or do things that are likely to achieve outcomes
7) Obtain more information to inform program improvement as needed (Focus Area III, Unit 2, synopsis)

Child welfare supervisors are uniquely positioned to intervene into the child welfare system. They are able to influence decisions made at both the administrative and the practice level. Additionally, research suggesting that good child welfare supervision is important in both retaining child welfare workers and in facilitating the development of skills necessary to do the job well, have contributed to an increased interest in improving the quality of child welfare supervision as a cost-effective means of intervening into the child welfare system.

Improving the Quality of Child Welfare Supervision

In response to increased performance expectations and in an effort to address workforce deficiencies, states are developing comprehensive curricula and training programs for child welfare supervisors designed to improve the quality of supervision that child welfare workers receive (Cyphers, 2003; US GAO, 2003; Zlotnik et al., 2005). In fact, 49 states included a plan to improve the training provided to supervisors in their Program Improvement Plans (PIP) submitted to the federal government in response to their CFSR. Of these states, 42 identified substantive changes to the content of their supervisory curricula as a key systems level intervention intended to improve their ability to meet federal performance standards in the future (Children's Defense Fund, 2006b). Examples include Missouri's role demonstration model of supervision training program, the mentoring model of supervisor professional development being incorporated in Arkansas, New York's clinical consultation model for supervisors and the University of Michigan and University of Alabama's joint Child Protection Service Casework

Supervision Project. Virginia, for the first time, instituted mandatory training for supervisors (VDSS, July, 2008).

This handful of supervisor training programs represents the range of strategies and conceptual models that are being used across the country in the effort to improve the quality of child welfare supervision. Missouri's role demonstration model of supervision includes training supervisor to model appropriate techniques for their staff and then to provide direct supervision of their staff's efforts to utilize new knowledge and skills (Sundet, Mermelstein, & Watt, 2003). In Arkansas, a training and mentoring approach is being used to enhance supervisory skills. Focal areas include five key components of supervision: structure, regularity, consistency, case orientation, and evaluation (Murphy & Goodson, 2007). Supervisory training incorporates interactional and educational components including parallel process and the provision of support through mentoring (Griffin & Shiell, 2003). In New York, Schools of Social Work were invited to contribute expert clinical consultation through regularly occurring case staffing meetings with child welfare supervisors (Strand & Badger, 2005). The Universities of Michigan and Alabama School of Social work have collaborated on a project to utilize learning labs to "promote evidence-based practice and to help supervisors understand their role in effecting organizational change" (Shackelford & Payne, 2003). The VDSS curriculum uses the Kadushin casework supervision model traditional to casework practice and incorporates a focus on the role of the supervisor in transfer of training for workers (J. Tondrowski, personal communication, September 16, 2008).

Limits to What is Known

From one perspective, the diversity of models, methods, and conceptual emphases found in these child welfare supervisory training programs are an ideal opportunity for comparative

evaluations of how best to impact the quality of child welfare client outcomes through supervisory best practices which encourage best practices in service delivery. From another perspective, efforts to evaluate the effectiveness of programs that rely on vastly disparate outcome measures may not be an effective way to compile generalizable knowledge for practice. The argument presented here is that, at least in part, research in the field of child welfare practice is limited because of a lack of standardized measures of supervisory practice.

Rationale and Purpose of Study

During fiscal year 2009, approximately 3.3 million allegations of child abuse involving 6 million children were made to Child Protection Services (CPS) agencies. Almost 63% of these reports resulted in either an investigation or an alternative response. Nearly 24% of the investigations resulted in a determination that one or more children were victims of abuse or neglect (HHS, 2010 c). More than 255,000 children entered foster care in FY 2009, while approximately 276,000 exited care. On September 30, 2009 there were 423,773 children in foster care in the US. Their average length of stay in foster care was 26.7 months. Of those children, 115,000 were waiting to be adopted. (HHS, 2010 a).

Federal funding, which is only a portion of the total amount of funds expended, for child welfare, foster care, and adoption activities under Title IV-B and IV-E of the Social Security Act for FY 2012 are estimated to approach $8 billion. This money will be matched by state and local dollars at rates varying from 50% to 25% depending on the funding stream. Seven billion dollars of the federal funding streams will be used to pay for mandatory programs: Foster Care (including room and board costs associated with placements for children of low income families in foster care, program staff [foster care workers], and administrative costs related to foster care placements), adoption assistance payment to adoptive families, guardianship assistance,

independent living services for older teens in foster care and tribal technical assistance. In the proposed FY 2012 discretionary programs budget, only adoption assistance, adoption opportunities, and domestic violence prevention programs are slated for modest increases in funding. All other program areas will retain level funding or face budget cuts (HHS, 2011).

While penalties imposed by the federal governments in response to poor performance in the CFSRs are intended to encourage states to invest in improvement efforts, resources are not readily available to substantially reform the states' child welfare systems. Furthermore, in these challenging economic times, there are real concerns that imposing penalties on already struggling states will translate to reductions in service provision, directly impacting the vulnerable children and families who are served by the child welfare system.

States are under significant pressure to address system and performance issues which are challenges to meeting the performance outcomes specified by the CFSR process. In order to accomplish these goals, states are undertaking significant efforts to improve the quality of their child welfare workforces and the services they provide through the child welfare system. In particular, efforts are being directed toward improving the quality of child welfare supervision in hopes of increasing staff retention and improving the skills of child welfare workers in a cost-effective fashion. However, these efforts are hindered by limited information about 1) the current practice of supervision in child welfare agencies, 2) which supervisory practices have the greatest influence on worker retention, 3) which supervisory practices have the greatest influence on improving service provision, and 4) what are the most effective interventions for improving the quality of child welfare supervision.

Research in these areas has been undertaken, but more information is needed. In order to improve the quality of child welfare supervision, the quality of child welfare services and,

ultimately, outcomes for children and families served by the system it is necessary to better understand how child welfare workers experience supervision. The present study is an attempt to increase understanding about the process of child welfare supervision through the development of a standardized instrument which measures supervision through analysis of supervisory behaviors from the supervisee's perspective. A standardized instrument will allow for comparison between studies and a generalization of findings in order to develop better supervisory training programs, to improve the quality of child welfare supervision in general, and to contribute to improvement of the quality of services provided to children and family served by the child welfare system.

Chapter 2: Literature Review

This review of literature is organized into four major areas of focus: 1) the child welfare workforce 2) social work supervision 3) supervision in child welfare and 4) empirical research regarding social work supervision and child welfare supervision. The relationship between the social work profession and child welfare practice is lengthy and interwoven. However, both social work and child welfare also have substantial bodies of supervision and workforce literature which are distinct. This literature review is an effort to present relevant information from both the social work and child welfare fields regarding supervision and child welfare workforce issues with an emphasis where these bodies of knowledge overlap.

The Child Welfare System

The child welfare system generally refers to those public agencies and their staff who are responsible for responding to reports of child endangerment, abuse, and neglect, and for intervening to protect vulnerable children. Interventions can include court action, including protective orders and removal of children from their parents' custody if necessary, to protect them. Child Protection workers perform assessments or investigations depending upon the severity of the reported endangerment and the policies of the state in which they practice. Children removed from their homes due to findings of abuse or neglect or through court action related to a caregivers' incapacity or failure to meet the child's special needs may be placed into foster care (Black, 2003; VDSS, 2008). Mental illness, substance abuse and dependence and

domestic violence are factors which greatly increase the risk of abuse or neglect (US GAO, 2006). Foster care placements can range from foster homes to acute psychiatric hospitalization depending on the assessed needs of the child. Increasingly, foster care has become an intervention that is used when older children engage in delinquent or out of control maladaptive behaviors or have extensive unmet treatment needs related to physical or mental illness or substance abuse (VDSS, 2007). Once a child is placed in foster care, his or her parents must prove that they are able to meet the child's needs safely before he or she will be allowed to return to the home (Black, 2003; VDSS, 2007). Child welfare workers are quite literally responsible for making life altering decisions regarding the needs and safety or risk of harm for children and the "fitness" of their parents.

Child Welfare Workforce

Child welfare jobs, which, at one time were held by the elite of the social work profession, have increasingly been filled by unskilled and untrained staff (American Public Human Services Association [APSHA], 2005; Schorr, 2000). The child welfare workforce of today is confronted with an increasing numbers of child abuse reports, increasing numbers of children in foster care, limited resources, and excessive caseloads comprised of increasingly complex cases. Worse, child welfare workers are inadequately prepared or supported in their efforts to serve multi-problem families and vulnerable children (Ellett & Leighninger, 2007; Tracy & Pine, 2000; US GAO, 2003). Turnover has been estimated to be between 30 and 40% annually nationwide. The average length of tenure for a child welfare worker is two years (US GAO, 2003).

Throughout this paper the term child welfare worker is used to refer to those individuals who work with children and families served by the child welfare system, including child

protection services and foster care workers. Although there are studies which suggest that BSW or MSW educated professional social workers are better qualified (Albers, Reilly, & Rittner, 1993; Booz, Allen, & Hamilton, 1987 as cited in Zlotnik, 2006; Ryan, Garnier, Zyphur & Zhai, 2006) and more committed to the field (NASW, 2004), only about one quarter of child welfare workers hold professional social work degrees (US GAO, 2003). Minimum educational requirements vary widely from state to state, jurisdiction to jurisdiction and agency to agency (Zlotnik, 2006). In 2000, the Child Welfare League of America [CWLA] found that only 28% of child welfare staff nationwide held either a BSW or an MSW and fewer than 15% of child welfare agencies required child welfare workers to hold these degrees (CWLA, 2001).

In their sample of 3877 child welfare workers from the state of Missouri, Ryan et al. (2006) found that only 7% had MSWs. In Georgia, Ellett, Ellis, Westbrook, and Dews (2006) found in their sample of child welfare workers from across the state, that 22% had a Master's degree, 63% had a Bachelor's degree and 4% had a high school diploma or GED. Majors most strongly represented included sociology (15%), social work (14%), psychology (9%), business (7%) and criminal justice (5%). Nearly 100 other majors were also noted including theology, political science, engineering, physical education, music, etc. More recently, in their sample drawn from a single large child welfare agency in the Western region of the country, Travis and Mor Barak (2010) found that 64% of the 359 respondents had Master's degrees. Of these, the largest group, 43% of the sample (n= 155), had MSWs. Thirty-four percent of the sample had Bachelor's degrees, though only 7.3% of the sample had a BSW (n= 26).

Frequently, the general public and mainstream media refer to child welfare workers as "social workers," regardless of the individual worker's educational background or training. This is especially discouraging when it occurs in a news stories concerning children's deaths. This

year in Washington, DC, there have been several child fatalities which have been well-publicized in *The Washington Post*. On February 2, 2008, a story regarding the death of Banita Jack's four daughters due to starvation included the statement by Mayor Adrian Fenty that six fired "social workers" who "just didn't do their jobs" (Stewart, 2008). A 6-month-old and a 4-month-old child died in separate incidents over the summer; Mayor Fenty again fired "social workers." In neither instance was any information provided as to the actual qualification of the individuals fired. On July 15, 2008, Stewart for *The Washington Post* reported that in the aftermath of these cases reports of child abuse have soared, while among remaining social workers "morale is so low that people often do not show up to work" (Dvorak, 2008).

Media attention of this type has a significant negative impact on child welfare workers in an agency where a death or other significant mistreatment of a child has occurred (Gustavsson & MacEachon, 2004; Steppe & Jones, 2007; US GAO, 2003). Additionally, negative press may contribute to a reluctance on the part of new BSW and MSW graduates to seek employment in the child welfare system due to concerns about a lack of professionalism in that setting (Ellett & Leighninger, 2007). Some leaders in the social work profession feel called upon to point out the differences between a professional social worker and a child welfare worker, particularly in the wake of unethical or particularly egregious errors being made public. Title protection legislation advocated by the National Association of Social Workers (NASW) at the state level can be seen as an attempt to distance professionally trained "social workers" from front line "child welfare workers" and other human services professionals performing in jobs that have traditionally been filled by social workers.

However, a more common response is the call to re-professionalize the field of child welfare. Support for this effort has been generated by educational institutions (Perry, 2004), the

NASW (NASW, 2004), the Child Welfare League of America (CWLA, 2001) and other federal and state agencies (US GAO, 2003; Zlotnik, 2003). More than 40 states have developed a Title IV-E funded training program to support child welfare workers pursuit of a BSW or MSW or to recruit promising BSW and MSW students who commit to working in the field post-graduation (APHSA, 2005; Zlotnik et al., 2005). Research suggests that Title IV-E trained child welfare workers are better prepared than their co-workers and tend to stay longer in their jobs (Fox, Miller, & Barbee, 2003; Jones, 2002; Jones & Okamura, 2000; Scannapieco & Connell-Carrick, 2003).

The field of child welfare is uniquely associated with the social work profession. It, more than any other practice area, has been developed and staffed primarily by social workers since its infancy (Lindsey, 2004; Schorr, 2000). In the minds of the general public, child welfare workers are social workers. It is in the best interest of the social work profession, therefore, to re-invest and recommit to efforts to improve service provision and outcomes for children and families. Moreover, the work done by the child welfare worker is social case work, arising from the traditional case work practice of the earliest child welfare workers who were, of course, pioneers in the field of social work (Woods & Hollis, 2000).

Social casework is defined as: " the orientation, value system and type of practice used by professional social workers in which psychosocial, behavioral, and systems concepts are translated into skills designed to help individual and families solve intrapsychic, interpersonal, socioeconomic, and environmental problems through direct face to face relationships" (Black, 2003, p. 402). Child welfare workers conduct complex family assessments, make decisions about individual and family needs and determine which services will be most effective. They make referrals. They monitor progress. They intervene directly in crisis situations and attempt

to increase the family's involvement or participation in services. They insure that clients are taking advantage of all of the resources that are available to them in the community. They address both concrete needs and interpersonal and psychological problems (Black, 2003; CWLA, 2001; Perry, 2006a). In short, the child welfare worker provides social case work services for the child and family, whether they are trained as professional social workers or not.

Although the work that is being done in child welfare agencies across the country continues to require the skill set of the professionally trained social worker, currently many child welfare workers are not professionally trained social workers (US GAO, 2003). As a result, the role of the child welfare supervisor in monitoring child welfare practice and in facilitating the development of worker skills is critical to the issue of workforce development and to the goal of improving the quality of services provided.

Social Work Supervision

Because child welfare supervision is so understudied, and because the field has been so closely linked historically with the field of social work, conceptual models of social work supervision are frequently used to describe the practice of supervision in child welfare (e.g. Cearley, 2004; Collins-Camargo, 2006; Faller, 2004; Griffin & Schiell, 2003; Jacquet et al., 2007; Jones, Alexander & Patterson, 2003; Perry, 2006b; Shackelford, & Payne, 2003; Smith, 2005; Strand & Badger, 2005, 2007). Social work supervision is a process used to help social workers further develop their skills, maintain positive morale and provide quality assurance for clients (Black, 2003; Kadushin & Harkness, 2002; Tsui, 2005).

Supervisory functions. Social work supervision is often discussed in terms of supervisory functions. While the number and content area of these functions can vary across models, three functions tend to be generally recognized in the literature: administrative,

educational and supportive. Administrative supervision is generally focused on agency policy and public accountability; supervisors coordinate, direct and monitor social work practice. Supportive supervision refers to the activities of the supervisors intended to assist the supervisee to develop and maintain reasonable expectations of him or herself, to promote the use of healthy coping strategies and to provide a sufficiently safe relationship within which the supervisee can express his or her feelings about the work (Kadushin & Harkness, 2002; Tsui, 2005). The focus of educational supervision is on professional development: developing self-awareness, and improving case specific knowledge and skills (Black, 2003; Kadushin & Harkness, 2002; Tsui, 2005).

Stages of the supervisory process. Across multiple supervision models, the supervisory process is also described as having stages or phases of development. Supervisees have different needs at different stages in the process; the supervisor's responses must be tailored to take this into account (Black, 2003; Tsui, 2005). Shulman (1993) identifies four phases of supervision: the preliminary stage, the beginning stage, the work stage, and the termination stage. In each stage, specific skills are required of the supervisor to facilitate the development of a productive supervisor- supervisee relationship. Kadushin and Harkness (2002) and Munson (2002) also address the importance of orienting the new supervisee to the supervision process. Generally, the new, less experienced supervisee is expected to need more didactic instruction and a tutorial supervision method (Black, 2003). With more experience, supervisee should become more independent and use supervision to focus on professional development goals and to manage work related stress and distress (Kadushin & Harkness, 2002; Tsui, 2005).

History of social work supervision. In the earliest history of the social work field, the Charity Organization Society of the late 1880s was largely concerned with training and directing

the efforts of the volunteer "visitors" who worked directly with families in need of assistance. This responsibility was primarily vested in the few paid "agents" of the organization (Tsui, 2005). Agents assigned cases and instructed visitors regarding their responsibilities. Volunteers were frequently unprepared for the work they were assigned and turnover was a significant problem. Supervision of the visitors is believed to have emphasized educational and training aspects as well as the provision of emotional support and encouragement. Although group supervision was utilized for training purposes, the principal context for supervision became the individual conference and the case record became the vehicle (Munson, 2002; Kadushin & Harkness, 2002).

With the advent of schools of social work in the early 1900s and the increased emphasis on professional training, educational supervision became the most common type of supervision (Tsui, 2005). The model of field instruction, where an inexperienced social worker was paired with a seasoned social worker became institutionalized as a social work supervision method (Munson, 2002). Primary responsibility for the provision of education was gradually transferred from the agency to the universities, but the agency supervisors retained responsibility for facilitating the transfer of educational learning into the field (Kadushin & Harkness, 2002). It wasn't until the mid-1960s that scholars and researchers began to recognize that agency supervision differs in significant ways from field instruction, primarily through its primary focus on the client, rather than on the worker (Tsui, 2005).

In the 1920s and 1930s, there were major changes in practice which in turn influenced the provision of supervision. Psychoanalytical theory was integrated into social work practice. The supervisory process began to be seen as a vehicle for facilitating "self-awareness" in the social worker. Supervision took on some aspects of the "therapeutic" process and in the 1950s the

concept of “parallel process” was incorporated into the supervisory model, describing the ways in which supervisees use the same skills to help their clients that their supervisors use to help them. During this period, the focus of supervision was on the professional development and growth of the supervisee (Munson, 2002; Tsui, 2005). Today, the emphasis on professional self-awareness and the personal and affective aspects of supervision which grew out of this period, continue to be relevant. However, examination of the personal life of the supervisee is no longer considered to be required or even an appropriate function of supervision (Tsui, 2005).

In the 1940s and 1950s, there was an increasing focus on establishing social work as a profession. Independent practice and commitment to ongoing learning were considered to be essential aspects of a profession. Ongoing supervision was seen as a barrier to achieving professional status. Advocates for independent professional practice pushed for the development of standards, which when met, meant that the social worker could practice without supervision (Munson, 2002; Tsui, 2005). Standards for independent social work practice have been institutionalized in state law and through professional organizations including the NASW, as licensure. Independent social workers are responsible for seeking out ongoing educational experiences and are encouraged to use consultation (rather than supervision) as needed (Tsui, 2005). Munson (2002) notes that credentialing can be seen as a substitute for supervision and, as such, can have a detrimental effect on social work practitioners as well as clients.

Tsui (2005) identifies the “growth of managerialism” in the 1980s as the last major trend to effect social work supervision. Managerialism refers to an increased focus on cost-effectiveness and management of resources. Social work staff are viewed more as employees than professionals. Funding bodies have increased authority to dictate provision of services. Tsui states, “Under managerialism, the emphasis is on job performance, task orientation,

standardization, documentation, consumerism and cost awareness. The worker-client relationship becomes a transaction instead of a transformation" (p. 8). This emphasis has resulted in an increased focus on administrative aspects of supervision—holding the worker accountable and promoting effective and efficient service to clients (Munson, 2002; Tsui, 2005).

Defining social work supervision. Tsui and Ho (1997) identify three approaches to conceptualizing social work supervision: 1) the normative approach 2) the empirical approach and 3) the pragmatic approach. The normative approach is described as asking key questions such as "What should supervision be?" It is the attempt to identify the norms or standards of supervision through its ideal functions. Tsui and Ho suggest that the normative approach focuses on the agreed upon objective of social work supervision: to improve the worker's capacity to do his/her job effectively by providing a good work context, knowledge and skills for professional development, and emotional support for staff morale.

The empirical approach focuses on collecting information about what supervisors actually do (Tsui, 2005; Tsui & Ho, 1997). Attempts are made to answer this question through collecting information about the roles and behavior of social work supervisors (e.g. Erera & Lazar, 1994a; Kadushin, 1992, Poertner & Rapp, 1983; Shulman, 1993).

The pragmatic approach sets aside formal definitions and focuses on "the functions and tasks of social work supervision in order to identify its nature and characteristics" (Tsui & Ho, 1997, p. 184). Functional definitions of supervision have largely focused on varying balances of administrative, educational, and supportive functions, and debates over which ranks first in the hierarchy. There have been continuing debates over whether administrative and educational functions are compatible. Some scholars argue that they are not only compatible, but also mutually beneficial (Kadushin & Harkness, 2002, Kadushin, 1992). Others argue that separating

the two achieves greater supervisory effectiveness (Erera & Lazar, 1994b). Furthermore, the pragmatic approach has yielded the interesting finding that there is a gap between what supervisors believe are the most important functions of supervision and what they actually spend their time doing (Kadushin, 1992; Poertner & Rapp, 1983).

Child Welfare Supervision

Supervision in child welfare is most often defined pragmatically, that is, by supervisor responsibilities or functions. For example, in the HHS's *Supervising Child Protective Services Caseworkers* manual, Salus (2004) states: "The primary function of child protective services (CPS) supervisors is to achieve the outcomes of safety, permanency and well-being for the children and families they serve and to enable casework staff to fulfill the mission of the agency" (p. 7). The Child Welfare League of America's (CWLA) *Standards of excellence* (1995) indicates that the foster care agency: "should provide [foster care workers] with regularly scheduled supervision with supervisors who have the competences to provide support; promote growth on the job; ensure that administrative and legal responsibilities are met; determine that performance standards are met; and provide individualized training" (p. 113-4). Training of new workers and overseeing the ongoing practice of established workers is seen as an essential role of the supervisor in the provision of quality services to clients (Salus, 2004). Child welfare supervision is seen as a process used to help child welfare workers further develop and refine their skills, to provide emotional support, and to monitor the provision of services. Frequently the role of the supervisor in promoting staff morale and job satisfaction through the provision of support is emphasized (US GAO, 2003, 2006).

Clinical supervision in child welfare. Clinical supervision in child welfare often refers to supervision which emphasizes the direct interaction or interventions provided by the social

worker to the clients (Munson, 2002; Salus, 2004). In the field of social work, "clinical social work practice" can be particularly distinguished from child welfare practice. In his introduction to his *Handbook of Clinical Social Work Supervision*, Munson (2002) writes that "agencies are making distinctions between case management practice and clinical practice" (p. 8). He specifically references child welfare practice as case management, which he distinguishes from clinical social work, which includes the direct provision of mental health services as opposed to making referrals for services. This view has certainly been adopted in the field of child welfare as well, where case management rather than psychotherapy is the primary service provided. There has been resistance within the field to recognizing the importance of offering clinical supervision (National Child Welfare Resource Center for Organizational Improvement, 2007).

However, child welfare practice is better characterized as case work than case management, and as such, has a significant clinical component (Woods & Hollis, 2000). The child welfare field is increasingly focused on efforts to improve the quality of services which are provided through improved decision-making, use of evidence based practices and better assessments, referrals and interventions by child welfare workers in the field. Public agencies are seeking to attract professional social workers to the workforce because they are better prepared to practice at the desired level (US GOA, 2006, 2007). Good child welfare practice requires child welfare workers to have good clinical skills. Therefore, clinical supervision an important part of child welfare supervision (Collins-Camargo, 2006; Jones et al., 2003; National Child Welfare Resource Center for Organizational Improvement, 2007; Strand & Badger, 2005, 2007).

Provision of child welfare supervision. There are five formats of supervision widely discussed in the supervision literature: individual supervision, group supervision, peer supervision, team management or consultation. (Kadushin & Harkness, 2002; Tsui, 2005). In

child welfare agencies, individual supervision is the most common format utilized, although, others, particularly group and team management, are occasionally used as well (Salus, 2004).

Individual supervision in child welfare has been greatly influenced by the practice model of social casework (Kadushin & Harkness, 2002; Tsui, 2005). The format of individual supervision very closely resembles the format of casework intervention. Provision of individual supervision consists of a one-on-one relationship where the supervisor and supervisee meet together, usually once a week, for an hour to two hours. The role of the supervisor encompasses the administrative, educational and supportive functions. This model is also very similar to the field instruction model and most supervisors use the one-on-one tutorial method they learned from their fieldwork supervisors (Tsui, 2005). Kadushin & Harkness (2002) and Tsui (2005) indicate that this is the most widely used model of supervision; this model is particularly useful with inexperienced workers and in settings where a high degree of administrative oversight is preferred.

The group supervision model is the second most widely used format of supervision in child welfare. It is often used as a supplement to rather than a substitute for, individual supervision (Kadushin, 1992). Group supervision uses the group setting to fulfill the responsibilities of supervision. Group members are exposed to a wide variety of learning experiences, and emotional support is provided by group members, as well as the supervisor. In group supervision, the supervisor's role is that of group leader, encouraging staff members to share their challenges and offer insights to each other. The objectives of group supervision are the same as those of individual supervision, however, the group format saves time and resources. On the other hand, group supervision may stimulate competition or relieve individual members of the responsibility to engage in problem solving. Critical feedback may be inhibiting to some

members and it is frequently difficult to integrate new staff persons into an existing group. The success of the group supervision model depends on the supervisor's ability to balance power between the leader and the group members, to create a safe space where workers are able to disclose dissatisfactions and concerns, and to insure that the needs of the individuals are not so diverse that some individual needs cannot be met or that competition develops in the group (Tsui, 2005; Watson, 1973).

In the team management format of supervision, the supervisor serves as the team leader. There are no supervisory sessions. Instead, the team uses group decision making to manage the work of the team. The team leader may have the final say, but each member is actively involved in decisions related to work assignment, performance monitoring and professional development (Kadushin & Harkness, 2002).

While the format of supervision offered to child welfare workers may be a function of agency resources, the needs of the staff, or simply tradition, the conceptual model of supervision utilized by the child welfare supervisor influences his or her emphasis and interventions regardless of whether he or she is providing individual or group supervision, or as part of team management.

Conceptual Models of Social Work Supervision

In order to understand what child welfare supervisors do to meet their responsibilities, more than a pragmatic description of duties is required. There are several conceptual models of social work supervision already being used to inform the development of training programs for child welfare supervisors, which are available for use in the empirical study of child welfare supervision. The models differ widely in terms of emphasis. Additionally, each, to some degree, borrows from the broader conceptual literature on supervision so that one or more functions of

supervision (e.g. supportive, administrative, etc.) are either incorporated into and contributing to other structural concepts or are central to the model. While each can be utilized in an individual or group format, in this discussion, individual supervision is addressed.

Interactional supervision. Shulman (1993) defines interactional supervision as a process in which the supervisor serves as a mediating factor interacting with various systems using four phases of work (preliminary, beginning, middle and end). His approach is based on three major tenets: 1) common dynamics and skills are central to all supervision processes; 2) these dynamics occur in all the different types of interactions in which supervisors are involved; and 3) there are parallels between the dynamics of supervision and other helping relationships. The supervisor-supervisee relationship provides a problem-solving model and a training opportunity for the supervisee. The supervisor- supervisee interactions inform the worker- client interactions.

In the preliminary stage, Shulman suggests that the most important skills for a supervisor is "tuning-in" or developing a well-developed sense of the new worker's background, values, culture, expertise and habits. The supervisor should develop a degree of empathy by taking the time to familiarize him or herself with the supervisee and his or her experience of entering into a new supervisory relationship and/or social work practice setting. The supervisor should not take skills for granted or have unreasonable expectations, instead issues should be addressed directly and non-judgmentally and free communication between the worker and the supervisor should be encouraged. It is important that the supervisor convey concern for the supervisee as a person, as this preliminary stage is the foundation for the supervision relationship. The goal for the supervisor of the preliminary stage is to sensitize him or herself to the fears and feelings which

may be present in the relationship but which the supervisee may not yet be comfortable communicating.

In the beginning stage, the focus of supervision should be on establishing mutual trust between the supervisee and supervisor and on developing a mutually agreed upon supervisory contract. While the contract can be a formal or informal agreement, it is essential that the negotiation process include open communication and, as a result, the supervisee begins to develop a sense of trust in the supervisor. New supervisees may have concerns or may not be clear what the purpose of supervision is. They may have questions about how authority and power may be used. They may worry that certain questions or concerns are inappropriate or reflective or significant performance problems, and so may avoid addressing them. A significant amount of anxiety on the part of both the supervisee and supervisor can be alleviated through a direct discussion regarding expectations and concerns. The supervisor should share with the worker his or her understanding of each others' roles and the purpose of supervision and mutual obligations and expectations. The supervisor should initiate this conversation but should be careful to elicit feedback from the supervisee. In addition to contracting, this stage is marked by the supervisee's "testing" what is acceptable and unacceptable. This phase concludes when mutual obligations and expectations are established.

Shulman (1993) identifies the work phase as the core supervisory stage in which the supervisor uses eight key skills: sessional tuning-in; sessional contracting ; elaboration skills; empathic skills; sharing skills; assertive skills; skills is sharing data; and sessional ending skills. The supervisor should utilize sessional tuning-in to notice problems and encourage the supervisee to address them directly in the session. Sessional contracting should be used to set an agenda for each conference. This both focuses the work and empowers supervisees to make the

conference time meaningful. Elaborating skills are used when the supervisee shares a particular concern. In many cases, listening may be more important than giving advice or information. Asking questions to develop a better sense of what the supervisee is thinking and feeling, will insure that the correct issue is responded to. The supervisor may need to consciously contain his or her own feelings, desire to help or anxieties in order to facilitate the supervisee's elaboration of the issue.

Empathic skills allow the supervisee to experience the supervisor as really understanding what he or she is feeling and that those feelings are normal and acceptable. This can be immensely reassuring, especially to a new worker. Empathic skills allow the supervisor to demonstrate sincerity and encourage a feeling of connection in the supervisory relationship. Skills in sharing feelings are a way to model feeling sharing for the supervisee. Rather than force the supervisee to share, supervisors can show vulnerability, acknowledge their own strengths and weaknesses and thereby facilitate the supervisee's comfort in also doing so.

Supervisors are also responsible for ensuring the quality of work of their supervisees. Assertive skills are important to help the supervisee recognize expectations and demands. Limit-setting may be necessary when demands are not met. Skills in pointing out obstacles encourage the supervisee to focus on job performance. Although there may be personal issues as well as organizational issues which are impairing performance, the supervisee's personality should not be the focus of attention. Instead skillful questioning regarding barriers to performance and the clear setting of expectations are required. The supervisor is also responsible for sharing information necessary to do the job well with his or her supervisees. Skills in data sharing refer to the process by which information is made available to supervisees in way that encourages them to ask questions and to use the information in problem-solving. Finally, in the work phase

of the supervisory process, the supervisor uses sessional ending skills to summarize what has been discussed and identify the next steps in achieving agreed upon goals.

The final phase of supervision is the ending phase. Similarly to the termination stage with clients in social work practice, the ending phase is an opportunity for the supervisor to summarize the various phases of the supervisory process and point out what the supervisee has learned or how he or she has grown. The supervisor can express appreciation for both the work the supervisee has done and for the effort he or she has invested in the supervisory process. Supervisors can model appropriate expression of feelings, including ambivalence, around leave-taking.

The Shuman model of interactional supervision has informed child welfare research studies as well as supervisory training programs (e.g. Cearley, 2004; Jones et al., 2003). The emphasis on parallel process and the importance of the supervisor-supervisee relationship fits well with the goal of intervening into the child welfare system through improving the quality of supervision provided.

Clinical social work supervision. Munson (2002) focuses primarily on three aspects of clinical supervision: 1) the supervisor-supervisee relationship; 2) the worker-client relationship; and 3) the administrative aspects of clinical supervision. Clinical supervision is defined as "an interactional process in which a supervisor has been assigned … to assist in and direct the practice of supervisees in the areas of teaching, administration, and helping. The supervisees are graduates of accredited schools of social work who are engaged in a practice that assists people to overcome physical, financial, social or psychological disruptions in functioning through individual, group, or family intervention methods" (p. 10). Supervisors' activities are organized into five major categories: 1) staying current with appropriate literature; 2) writing-including

reports, grants and articles; 3) observing the job performance of their supervisees; 4) listening; and 5) providing information and facilitating discussion.

Munson emphasizes the importance of the clinical supervisor's role in helping new MSW graduates to acquire advanced clinical skills. He identifies the following trends as increasing the need for clinical supervision: "the resurgence of clinical [social work] practice; social work dominance in mental illness treatment; the theory explosion; changes in professional standards, limited supervision training, practitioner stress and burnout prevention; unethical practice; complex external controls on practice structure; complicated documentation; sophisticated outcome measures; and new highly specialized practice areas" (p. 23). The recent graduate needs assistance to integrate practice demands with what he or she has been exposed to through the educational process. Munson states strongly that the MSW does not adequately prepare the graduate for independent clinical practice, and that the primary purpose of clinical supervision is "to guide the supervisor in efforts to enhance and refine the functioning of the recent graduate" (p. 23).

The learning needs of graduating MSW include: "exposure to specialized cases; criticism of work; exposure to different practice approaches, more exposure to theory, co-therapy experience; exposure to the work of others; more direct supervision; developing more self-awareness; feedback on work; support and encouragement; training in group therapy; and help in improving diagnostic skills" (p. 36). Clinical supervisors, then, are both teachers and role models. In particular, they facilitate the supervisee's ability to productively conceptualize a case, to recognize patterns of behavior, to utilize theory effectively in developing treatment plans, and to feel confident in his or her ability to intervene.

Munson (2002) argues that the power of the supervisor should come from the supervisor's competence, that is, his or her ability to do the job, rather than from his or her position. The competency model of authority has been found to be significantly correlated with supervision satisfaction and job satisfaction. The supervisor should have wisdom and experience to pass on, as well as a commitment to always growing and improving as a practitioner. The supervisor has an important role in transmitting the mission and vision of our social work heritage, and in inspiring the supervisee to take on this shared mission with skill and passion.

Although Munson suggests that clinical supervision is only appropriate in settings where "clinical" work is being performed by professionally trained clinical social workers, there are elements of this model that have been incorporated into child welfare supervisory training programs (e.g. Collins-Camargo, 2006; Strand & Badger, 2005, 2007). As MSWs are recruited into the child welfare workforce, there is an increased need to provide them with opportunities for professional development; clinical supervision towards licensure in one means of doing this. Perhaps more importantly however, the provision of clinical supervision to child welfare workers emphasizes the importance of using clinical skills in decision-making and intervention in child welfare practice.

Comprehensive model of social work supervision. Current work on both social work supervision and child welfare workforce and practice issues research suggests that organizational climate plays an important role in client outcomes (Glisson & Hemmelgarn, 1998) and staff retention (Bednar, 2003; Wagner, van Reyk, & Spence, 2001). Tsui (2005) has proposed a new social work supervision model which incorporates the interrelationships between four parties: the agency, the supervisor, the staff, and the client. This model is strongly derivative of the models developed by Munson, Shulman and Kadushin (see below), but incorporates "culture," as the

context in which the four parties interact with each other as an important variable. Tsui suggests that this comprehensive model is an important "widening" of the perspective of supervision consisting solely of supervisor-supervisee interactions to include organizational appropriateness, contextual specificity, and cultural sensitivity, in addition to supervisor-supervisee interactions.

Tsui also suggests that research which assesses the context of supervision as well as the supervisor-supervisee relationship is necessary to better understand the process of supervision. Additionally, efforts to improve supervision would necessarily include efforts to address the physical, interpersonal, cultural and psychological environment in which supervision occurs. The physical environment refers to the actual space or venue where the supervision session is held. The interpersonal context refers to the dynamics between the supervisor and supervisee. The cultural context refers to the norms and values of the society in which the supervisor and supervisee are practicing. Finally, the psychological context addresses the attitudes, beliefs and feelings the supervisor and supervisee bring to their sessions. Tsui states: "In this model, the effectiveness of supervision depends on several factors: the relationships among the individual parties (the agency, the supervisor, the staff, and the client); the contract, format, and developmental stages of supervisory process; the balance among the various supervisory functions; and the relationship between the features of supervision and the culture of the external environment" (p. 47).

According to Tsui, this conceptualization of supervision will lead to more comprehensive research including 1) the relationship between supervisory practices, job performance and client outcomes 2) the impact of specific cultural contexts on supervision and 3) the relationship between work environment variables (climate, organizational goals, organizational norms, etc) and the effectiveness of supervision. In the field of child welfare, supervision and organizational

factors have been studied separately and as related variables rather than in an integrated model as Tsui proposes. Unfortunately, as Tsui acknowledges, it is challenging to operationalize "culture." The comprehensive model is also very complex, with as many as 14 variables playing key roles. Tsui's point is well-taken; focusing solely on the supervisor-supervisee relationship neglects important aspects of the supervisory process. However, this model is largely untested and there are no existing procedures for doing so.

Casework supervision model. The casework model of supervision focuses on the three predominant functions of supervision: administrative, educational, and supportive supervision. A supervisor who is skilled in all three functions is a "good" supervisor. Kadushin and Harkness (2002) identify key responsibilities and roles of the supervisor with an emphasis on the importance of the supervisor-supervisee relationship. They provide the following definition of social work supervision:

> A social work supervisor is an agency administrative-staff member to whom authority is delegated to direct, coordinate, enhance, and evaluate the on-the-job performance of the supervisees for whose work he or she is held accountable. In implementing this responsibility, the supervisor performs administrative, educational, and supportive functions in interaction with the supervisee in the context of a positive relationship. The supervisor's ultimate objective is to deliver to agency clients the best possible service … in accordance with agency policies and procedures. (p. 23)

What distinguishes Kadushin's model from other models of social work supervision is its focus on the functions of supervision, which are defined explicitly and behaviorally. Moreover, Kadushin is interested in agency based social work supervision and practice and, therefore, both

studied supervision in child welfare agencies and directly addresses in his normative definition of supervision, how supervision "should be" in child welfare agencies.

Selection of supervision model for study. The goal of this study is to develop an instrument for use in studying the behaviors of child welfare supervisors from the perspective their supervisees. For the purposes of this study, a model of social work supervision focused on agency based or case work supervision, rather than clinical supervision, is more appropriate. As noted previously, although the practice model of the child welfare system is inherently based in clinical practices, it is not focused on psychotherapy. Therefore, of primary interest for this study are definitions and models of supervision which are suitable for case work supervision rather than models of clinical supervision or case management supervision.

Additionally, while the importance of organizational climate as a factor which influences supervisory behavior, client outcomes, and child welfare workers' job satisfaction is not denied, the focus here is on the child welfare workers' experiences of supervision. It seems important that organizational climate or culture be considered as a variable separate from supervision in studies of child welfare workforce issues as there is still much to learn about both. Indeed there have been several notable efforts to do just this (e.g. Ellett, Ellett & Rugutt, 2003; Glisson & Hemmelgarn, 1988).

The social work supervisor is uniquely situated in the social work agency to intervene both with the management above him or her and with the supervisee below. Indeed, being a conduit between the two and advocate for both, as appropriate, is an important administrative and supportive function of social work supervision (Kadushin & Harkness, 2002). It is assumed that the proposed measure will incorporate the supervisee's experience of his or her supervisor's intervention or advocacy efforts through assessment of the supervisor's behaviors. Furthermore,

there is a current need to assess the effectiveness of supervisory training programs which would benefit from a measure of supervisors' behavior separate from organization climate or culture. Tsui (2005) suggests that the comprehensive social work supervision model will contribute to theory generation, and that may well be the case. At this time, however, the model's complexity and relative newness are important factors in looking elsewhere.

Finally, the casework model of supervision allows for a focus on the development of clinical skills, although within this model it would be considered part of the supervisors' provision of educational and supportive supervision. It may be possible, then, to utilize workers' perceptions of the child welfare supervisors' educational and supportive behaviors to evaluate the impact of a supervisory training model focused on improving supervisors' clinical skills. Unlike the conceptual models of supervision that exclude a case management setting, the casework model was developed through studies of actual casework practice.

For the purposes of this study, supervisors' behaviors as perceived by the child welfare workers they supervise are of primary interest. The goal is to develop a behaviorally based measurement of supervision. Kadushin's model of casework supervision has been selected as the primary conceptual model to provide a framework for the development of the Experience of Supervision scale due to its explicit focus on supervisory functions (and the related activities), its appropriateness for use in settings where case management is the primary practice activity, its inclusivity and its simplicity in structure.

Casework Supervision Model and Child Welfare Supervision

Administrative functions. In the child welfare field, administrative supervision includes monitoring the case work of child welfare workers to insure that time lines and financial procedures are consistently followed. Other administrative functions include managing

caseloads/ case assignments, administering performance evaluations and addressing performance problems. In some child welfare agencies where supervisors have little clinical training or where is little importance placed on educational supervision, the primary focus of supervision may become administrative.

While providing administrative supervision, the supervisor can convey that learning and improving case management practice is a performance expectation through goal setting (including post-training behavioral goals), careful consideration of caseload when making case assignments (including protecting new workers or workers applying new work skills to their practice) and developing performance evaluations which reward participation in training and using new skills on the job (Kadushin & Harkness, 2002; Wehrman et al., 2002). Additional administrative functions include middle-management activities such as understanding, supporting and communicating the goals of the larger agency to the child welfare staff and also identifying and advocating with upper management for strategic efforts to improve systemic and individual worker effectiveness (Kadushin & Harkness, 2002).

Supportive supervision. Kadushin and Harkness (2002) define supportive supervision as "concerned with helping the supervisee deal with job-related stress and developing attitudes and feelings conducive to the best job performance" (p. 277). Job stress is generated by work with difficult clients, ambiguous or conflicting roles and responsibilities, constant change, and unreasonable personal expectations about job performance. Good supportive supervision effectively mediates between the structural stresses of the work environment and the supervisees (Kadushin & Harkness, 2002; Smith, 2005; US GOA, 2003). In providing supportive supervision, the supervisor is available to the worker for the ventilation of feelings, to provide positive feedback, and to hold out reasonable expectations in order to reduce, ameliorate and

prevent stress. The provision of supportive supervision enables the worker to mobilize the emotional energy needed for effective job performance and to obtain satisfaction in doing their job.

Additionally, supervisees receive support from a variety of sources other than the supervisor, but the supervisor can be instrumental in maximizing the availability and quality of those additional sources of support. For example, peer support is seen as a significant supportive resource available to child welfare workers. The supervisor can take an active role in facilitating positive peer interaction through leading regular staff meetings and/or group supervision, developing peer supervision and consultation opportunities, and structuring mentoring relationships (Kadushin & Harkness, 2002; Wehrman et al., 2002).

Finally, supervisors may initiate efforts to address negative organizational factors as a component of supportive supervision. In addition to facilitating the worker's healthy response to stress related to organizational context, the supervisor can actively engage in the process of trying to improve the organizational context through facilitation of better working relationships across teams or better interagency relationships, for example. The supervisor can advocate for reduced caseloads, recognition for outstanding performance (Kadushin & Harkness, 2002), and policies which support a specific multi-level response to crises for the staff (e.g. death of a client or worker being threatened of assaulted) (Gustavsson & MacEachron, 2004).

Educational functions. The chief function of educational supervision is the enhancement of the supervisee's practice skills through direct provision of knowledge and skills directly or by making training available and supporting gains made. The supervisor facilitates the application of new knowledge and skills to ongoing case practice through case consultation. In order to be effective, the supervisor must be proficient in the use of best-practice concepts and

skills. Additionally, the supervisor engages in continual assessment of each supervisee's current skill level and works with the supervisee to develop an individualized learning plan which will facilitate improved practice and professional development. (Kadushin & Harkness, 2002).

In his or her educational capacity, the supervisor can engage in effective assessment of the supervisee's learning needs and can strongly advocate for participation only in training which has a reasonable potential to enhance the particular supervisee's practice. Additionally, the supervisor can use this assessment (when appropriate) to frame the supervisee's attendance and transfer of learning as an "opportunity" rather than a requirement, further increasing the likelihood of training transfer (Salas & Cannon-Bowers, 2001; Wehrmann et al., 2002).

Finally, the supervisor is in a position to advocate for particular in-service trainings to be brought to the agency based on his or her unit's assessed needs and/or interests. The supervisor can also play an active role in advocating with administration for increased training funds and agency policy and practices which support educational goals, innovation, and worker self-efficacy (Kadushin & Harkness, 2002).

Empirical Studies

There are significant barriers to studying supervision in all areas of social work practice (Kadushin & Harkness, 2002; Tsui, 2005). Empirical studies of social work supervision practices in general are rare (Tsui, 1997, 2005), and of child welfare supervision in particular, more rare (Zlotnik et al., 2005). Ideally empirical studies of supervision would evaluate the effect of the supervisor's behavior on client outcomes as well as on the worker (Tsui, 1997).

Tsui (2005) suggests that there are two key factors which make research on social work supervision challenging. The first is that staff supervision, unlike field instruction supervision, is embedded in an organizational context. It is difficult to separate supervision from other

organizational factors which contribute to both worker attitudes and their practice with clients. Secondly, the power differential between the supervisor and the supervisee, the importance of performance evaluation to the worker's success in the agency, and the confidential nature of the supervisory process make it difficult to recruit supervisee and supervisor research participants.

The review of empirical studies of child welfare supervision will be focused in four areas 1) larger studies of social work supervision in which public agency based supervision is considered 2) evaluation of child welfare supervisory training programs 3) studies of child welfare workforce issues, including retention which include supervision as a variable of interest and 4) studies which attempt to link supervisory behaviors to improved practice and/or client outcomes.

Social work supervision. Many of the largest studies on social work supervision were introductory and exploratory in nature. Kadushin (1974, 1992) conducted two large scale national surveys on social work supervision in the United States in 1973 and 1989. He completed a basic analysis of the descriptive data and also identified strengths and weaknesses noted by supervisors and supervisees (Kadushin, 1992). Munson (1976) conducted an empirical study on the uses of structural, authoritative and teaching models using 65 dyads of social work supervisors and supervisees. Shulman, Robinson and Luckj (1981) conducted a large national survey of the context and the skills of social work supervision in Canada.

The different conceptual models of supervision developed by the authors of these three exploratory studies have become the foundational work on which subsequent studies have been based. For example, Erera and Lazar (1994b) developed a scale which operationalizes Kadushin's model of supervision from the supervisor's perspective. They note that the main difficulty in connecting results from different studies of social work supervision is the failure of

most studies to examine all of the supervisory functions. Their Supervisory Functions Inventory (SFI) consists of 39 action items such as "applying policies to working plans" (an administrative action item), "discussing assessment of clients with workers" (an educational action item) and "assisting workers with handling their feelings" (a supportive action item). The SFI was completed by supervisors, who were asked to rate the frequency of each action based on a 5 point Likert scale.

Erera and Lazar administered their instrument to supervisors in three types of agencies: social services departments, service-oriented agencies and treatment-oriented agencies. They were interested in looking at all the dimensions of supervisory function and whether there were differences between supervisors who performed both administrative and educational functions and supervisors whose responsibilities were mainly educational. Factor analysis suggested the factor structure composed of 7 factors, which were then grouped by theoretical dimensions. Three factors were classified as administrative, three as educational and one as supportive. Furthermore, the SFI distinguished between supervisors practicing in different agencies on the basis of administrative function. Department of Social Services supervisors were more likely to perform administrative actions than their counterparts in other settings.

A second finding of this study was that supervisors who performed more administrative functions tended to experience greater role conflict and ambiguity. Erera and Lazar concluded that educational and administrative supervision should be separated because the two roles are not compatible. Another conclusion might be that while supervisors with mixed functional responsibilities still tend to value educational supervision more (Kadushin, 1992), they spend most of their time performing administrative functions (Collins-Camargo & Groeber, 2003; Erara & Lazar, 1994a; Poertner & Rapp, 1983).

Limitations of the study include that all of the participants were from one country, Israel, and, therefore, the results may reflect a cultural influence that is not captured in the findings. Secondly, the SFI is based on supervisor self-report rather than task analysis which may limit the accuracy of the results. Finally, Erera and Lazar (1994b) operationalized Kadushin's supervision model from the perspective of the supervisor, but in the child welfare field the workers' perceptions of supervision are increasingly the focus of study (See Dickinson & Perry, 2002; Jacquet et al., 2007; Landsman, 2001; Mor Barak, Nissly, & Levin, 2001; Mor Barak, Levin & Nissly, 2006; Renner et al., 2009; Smith, 2005; Travis & Mor Barak, 2010; Zlotnik et al., 2005).

In another supervision study, Scott and Farrow (1993) attempted to evaluate the degree to which social work supervision being provided in child welfare and hospital social work agencies met the Recommended Standards for Social Work Supervision developed by the Australian Association of Social Work. These standards include that the supervisor be a professionally trained social worker, with at least three years of practice experience who has successfully completed an approved course in social work supervision. Other standards included frequency of supervision and that "supervision would meet the needs of the supervisee within the organization for administrative, educational, and supportive supervision" (p. 34).

The three functions of supervision were broken down into four specific activities, for a total of 12 supervisory behaviors. Each behavior was operationalized into two statements for a total of 24 items to which the study participants responded using a six point Likert scale indicating the degree to which the supervision they gave or received corresponded to the statement. Items included: "my supervision sessions include planning and assignment of work (administrative function)"; "supervision encourages me to expand my knowledge base

(educational function)"; and "supervision reinforces a positive sense of worth as a professional (supportive function)."

An eight-page questionnaire was administered to the entire population of social workers in Melbourne teaching hospitals and to a multi-stage cluster sample of child welfare supervisees and supervisors in Community Services Victoria, stratified by region and service delivery program, for a total of 233 respondents. Scott and Farrow found that hospital supervisors reported a slightly greater perception of educational function while child welfare supervisors indentified administration and supportive as more prevalent in their supervision. Among child welfare supervisors, the emphasis on administrative functions was more pronounced for those who were not social work trained.

Importantly, supervisees in both groups tended to perceive the supervision they received somewhat differently than the supervisors did. Additionally, hospital social workers perceived the supervision they received to be more heavily weighted towards supportive functions than did child welfare workers.

A major limitation to this study is that Scott and Farrow do not report any attempts to evaluate the reliability or validity of the instrument that they used. Furthermore, they do not report any statistical analysis of their findings beyond the report of mean values and therefore it is not possible to assess the statistical significance of the results.

Effectiveness of child welfare supervisor training programs. In response to the literature which overwhelmingly suggests that supervision plays an important role in preventing turnover and to the underlying assumption that improved supervision is key to the provision of improved services to clients, there have been a number of demonstration projects approved and

funded by the Children's Bureau within the US Department of Health and Human Resources specifically addressing child welfare supervisory training practices.

The diversity of supervisory training programs, in and of itself, suggests that best-practices in the area of child welfare continues to be based on limited empirical evidence and guided by diverse theoretical perspectives. For example, the University of Michigan School of Social Work developed a three-day training for child welfare supervisors. This training focuses on federal child welfare policy and best practices. Additionally, they developed a monthly half-day training program which is focused on management skills including "training and coaching staff, supportive communication, managing difficult people, problem solving and reviewing and evaluating staff" (Faller, 2004, abstract).

Arkansas developed a very different model of supervisory training based on mentoring. This model incorporates "interactional and educational components" drawn from the theoretical concept of parallel process. Strengthening the supervisor-worker relationship is intended to have a beneficial effect on the worker-client relationship. Specific techniques are taught with "the goal of improving the supervisor's knowledge and ability to enhance the supervisees' assessment skills" (Griffin & Shiell, 2003, abstract). Mentoring provides ongoing support for addressing supervisor-worker relationship issues.

A third model of supervision training can be found in the clinical consultation model developed in New York. Strand and Badger (2005) have reported on a three year, multidisciplinary clinical consultation model that was tested with child welfare supervisors in a large urban area. One goal of the model was to link MSW-level child welfare supervisors to faculty from schools of social work. It was found to "yield encouraging results for professional decision-making in the provision of direct service" (Strand & Badger, 2005, p. 2).

For each of the examples provided, program evaluation was undertaken to assess the effectiveness of the model. Each, to some degree, has been demonstrated to be effective: in comparing pre-test and post test scores on an assessment of knowledge, supervisors who participated in the Michigan supervisory training program showed significant improvement in factual information especially related to policy and best-practices after training (Faller, 2004); in self-reporting measures, supervisors who participated in Arkansas' mentoring model indicated that their supervisory skills had improved (Griffin & Shiell, 2003); and, supervisors who participated in New York's clinical consultation model reported increased job satisfaction and improved clinical decision-making skills (Strand & Badger, 2005).

Unfortunately, to date, none of these program evaluation efforts have included evaluation of whether training resulted in changes in the supervisors' actual behaviors, in a perception on the part of the supervisee that their supervision had improved, or in a change or improvement in the services provided by the supervisees, or in outcomes for children and families. A notable exception is a set of supervisory training initiatives funded in 2001 by the Children Bureau branch of the HHS and implemented through the Southern Regional Quality Improvement Center. In 2002, four states began 3 year demonstration projects designed to bring clinical casework supervision practices into the child welfare agencies. This intervention was hypothesized to positively impact worker satisfaction with supervision, foster more productive organizational climates within the agencies, reduce worker turnover, improve worker practice and have a positive impact on outcomes for children and families. Because the project was a public agency/university/community partnership, statewide data regarding many of these variables was more readily available and a large multisite evaluation was considered feasible (Collins-Camargo, 2006; Jones et al., 2007; Steppe & Jones, 2007).

While there were some differences in the intervention models and the addition of other variables of interest in some cases, the data collected through 1) state-wide data reports, 2) annually conducted worker and supervisor surveys, and 3) focus groups and case review regarding worker practice of both intervention and comparison groups were considered comparable. There were significant findings across the projects and at various points during the three years it was underway. Specific results for the Tennessee demonstration project (Jones et al., 2007; Steppe & Jones, 2007) will be commented on in greater depth below. Unfortunately, there were also significant problems which arose in evaluating the actual impact of the demonstration projects.

Collins-Camargo (2006) summarized all four states' projects' findings. The total sample size was not reported; the grant required that each state develop an intervention group of at least 20 front-line supervisors and a similar comparison group. The project specific results are described elsewhere. In two states, standardized measures of organizational culture yielded a significant improvement in organizational culture in intervention sites. Contamination between intervention sites and comparison sites was noted as a significant challenge to evaluation. Focus groups conducted with supervisors suggested that there was a perceived shift in focus from administrative supervisory activities to clinical supervisory activities and a shift in culture which enhanced self-reflective practice, and use of theory and research in practice.

Whether worker turnover decreased as a result of the supervisory training programs was not clear. Aggregate state level data was not detailed enough to permit comparison between intervention and comparison groups. The use of a standardized measure to assess intent to remain employed did not yield significant results.

In order to assess the degree to which worker practice was enhanced, evaluation focused on a measure of worker self-efficacy. This has been used in the literature as a proximal indicator of practice, because the degree to which an individual believes his or her actions will result in the achievement of a desired goal, is strongly related to the likelihood of him or her engaging in that behavior. Two of the states found statistically significant improvements in worker self-efficacy. Case records proved an ineffective means of evaluating actual changes in worker practice. Collins-Camargo (2006) reports: "overall the states involved reported a decrease in the achievement of practice standards over time in both intervention and comparison groups (p. 104)." During the three years of the project, the states were struggling to respond to deficits identified in their CFSR, budget constraints and leadership changes. It was not possible to tease out whether the supervisory projects had had no impact, a negative impact or a positive impact due to other conditions contributing negatively to worker practice.

Finally, there were concerns, even at the outset that it would be difficult to conclude what impact the project had had on outcomes for children and families, as three years is actually a relatively short period of time in which to see change statewide. Statistical testing was not utilized as the effect size was not predicted to be observable. Instead quarterly "pictures" of aggregate data were charted. In two states, the observable trends suggested that the supervisory training program had had a positive impact. Focus groups were also utilized for data collection. Collins-Camargo states: At the end of the intervention… supervisors in three of four states could provide very positive examples, including clients self-initiating treatment, increased engagement in case planning and even anecdotally based belief that cases were moving more quickly, and that children were being maintained at home or reunified more quickly" (p. 105).

Ultimately, the most significant evidence that the demonstration projects were successful were supervisor self-reports. This is seen less as an indicator that the supervisory training programs were not successful, and more as an indicator that the strategies for assessing success were perhaps not specific enough. Furthermore, the problems noted regarding the impact of other changes to the system is a common one when conducting any type of longitudinal study in a child welfare agency. Policy changes over the last decade have been dramatic. Budget issues, demoralizing scandals and sweeping changes in leadership are all relatively common occurrences, which can have significant affects which compromise ongoing research (Collins-Camargo, 2006; Steppe & Jones, 2007).

Child welfare supervision and workforce issues. Supervision has consistently been found to play an important role in the retention of child welfare workers. Child welfare studies have found that poor supervision is a primary reason for worker turnover (Samantrai, 1992; US GOA, 2003), while good supervision is a reason to stay (Dickinson & Perry, 2002; Mor Barak et al., 2006; Rycraft, 1994). Smith (2005) found that one standard deviation increase in reported level of supervision support increased the odds of worker retention by 46%.

Child welfare supervision has also been found to play an important indirect role in worker retention; supervisory support has been found to have a positive effect on intent to stay through its influence on job satisfaction and organizational commitment (Landsman, 2001). Supportive supervision has also been found to have a mediating effect on a number of other factors linked to turn over: job stress (US GAO, 2003), vicarious traumatization (Bride, Jones & MacMaster, 2007; Pearlman & Saakvitne, 1995), and burnout (Drake & Yadama, 1996; Kim, 2011). Furthermore, supportive supervision has been found to contribute significantly to factors

positively correlated with intent to stay: organizational climate (Bell, Kulokarni & Dalton, 2003) and work-life balance (Smith, 2005).

While studies regarding retention and turnover have contributed to our understanding of the relationship between supervision and other variables including retention and turnover, they have not yielded as much information about supervision as might be desirable. For example, in her 2001 study on commitment in public child welfare, Landsman used a social support index adapted from Caplan, Cobb, and French (1975, as cited in Landsman, 2001) in which supervision is one of three types of support measures. Supervisory support and co-worker support measures were taken directly from Caplan's measure and a related measure of agency support measures was developed for the study. Each subscale was comprised of only 2 or 3 items.

Landsman's findings suggest that job satisfaction, commitment and intent to stay in the profession and commitment to and intention to stay with the agency are distinct constructs. Furthermore, the degree to which employees are satisfied with their jobs affects the degree of which they feel positively towards the agency with whom they are employed. The only two variables in the model being tested which had substantial effects on job satisfaction were supervisory support and opportunities for advancement. In Landsman's model, supervisory support is an important component of job satisfaction, which has a strong relationship to intent to remain employed in the agency; this finding is frequently cited as evidence that supervisory support is a key strategy for reducing turnover.

However, Caplan's Social Support Index (which includes supervisor, co-worker and significant other/spouse subscales) was originally developed to assess the degree to which social support influences individuals' attempts to stop smoking. In Landsman's study, supervisor support is operationally defined as "the extent to which the employee believes the immediate

supervisor provides instrumental (knowledge or skill-based) and affective (emotional) support" (p. 409). Using Kadushin's description of supervisory functions, instrumental support may be more closely related to educational supervision functions than supportive supervision functions. Landsman does not provide the items, so it is not clear how these two types of "support" are represented in the scale. As a result, the finding that supervision support positively affects job satisfaction is less clear than it seems at first.

Mor Barak et al., (2006), also used Caplan's (as cited in Mor Barak et al., 2006) Social Support Index in their study on causes of preventable worker turnover. Their model included the supervisory support and co-worker support, but also included spouse/significant other support. Agency support was not a variable. Data was collected by means of a questionnaire administered to an availability sample of 418 participants drawn from a large urban public agency.

Mor Barak et al. found that the strongest predictors of intent to leave were lack of job satisfaction, low organizational commitment, younger age, high stress and exclusion from the organizational decision making process. A secondary finding was that workers who perceived more supervisory support also reported higher levels of inclusion in the decision making process. Additionally, workers who report more supervisor and more peer support also reported less stress. Analysis of the qualitative data collected through follow-up in depth interviews with 33 of the original participants supported Mor Barak et al.'s conclusion that supportive supervision plays an important role in the worker having a stronger sense of inclusion in the organization and can have a moderating effect on stress. Again, however, because of the way supervisory support was operationalized, the study does not help answer questions about what specific supervisory behaviors are experienced as supportive by the workers or whether a behavior which might (in another instrument) be classified as educational, is actually the factor contributing to the results.

As part of a recurring evaluation study of specially trained child welfare workers, Jacquet et al. (2007) studied the relationship between supportive supervision and intent to leave. In California, the California Social Work Education Center (CalSWEC) consortium specifically recruits MSW candidates to receive specialized training and financial support in return for a commitment to work as a child welfare worker for a specified period of time post graduation. The program is federally funded using IV-E. From May 1996 through April 2005, follow-up surveys have been sent to a total of 1498 CalSWEC MSWs who have completed their "pay-back" through work in public agency child welfare positions. The follow-up survey is part of the long term effort to evaluate the effectiveness of the California IV-E stipend program. However, information about other factors which have been found to influence retention is also collected.

Jacquet et al. found that among support systems, supervisor support was more predictive of intent to stay than peer or family support. They further examined the participants' perceptions of supervisory support through their responses to the following statements: my supervisor 1) …is competent in doing his/her job; 2) …is very concerned about the welfare of those under his/her; 3) …gives information when I need it; 4) … shows approval when I have done well; 5) … is willing to help me complete difficult tasks; and 6) …is warm and friendly when I have problems. Jacquet et al. found that for a single unit rating of positive regard for their supervisor, respondents were 1 ½ times more likely to stay employed in their child welfare agency. Respondents who indicated higher levels of supervisor support also indicated that they would have been more likely to stay, even if they hadn't been contractually bound. Furthermore, supportive supervision was found to be more predictive of intent to stay than caseload size, which has long been considered to be a key factor influencing retention.

Six behaviorally based questions about supervision received were used by Jaquet et al. as their measure of "attitude towards supervision." These items include educational and administrative as well as supportive functions, but are defined as "supportive supervision" in the analysis. As a whole, the composite measure or "supportive supervision" was found to contribute significantly to the model predicting retention, but little is known about how specific supervisory behaviors contributed to the overall "attitude towards supervision."

As a component of a larger study of the impact of a supervisory training program implemented by the Southern Regional Quality Improvement Center (mentioned above), Bride et al. (2007) assessed the relationship between Child Protection workers level of secondary traumatic stress and their personal histories of trauma, co-worker and administrative support, intent to remain employed in child welfare, professional experience, and size of caseload. A web based survey was used to collect data. A total of 187 CPS workers participated from across the state of Tennessee.

Perhaps the most significant finding of this study was that 92% of the participants reported experience at least one symptom of Secondary Traumatic Stress (STS) at least occasionally in the previous week. Fifty-nine percent reported experiencing one or more symptom often during the preceding week. A total of 34% of the participants met the clinical criteria for work-related PTSD as evidenced by their experiencing intrusions, avoidance and arousal symptoms often in the previous week. The study also yielded statistically significant correlations between STS and peer support, caseload size, personal history of trauma, and intent to remain. Administrative support was not found to be significantly correlated. However, peer support and case load size are factors which can be directly impacted by supervisory behaviors.

It is not clear to what degree supervisory behaviors were, if at all, measured in the study. The variable Administrative Support was measured by use of the Professional Organizational Culture Questionnaire-Social Work (POCQ-SW) (Ellett & Millar, 2001, as cited in Bride et al., 2007). This instrument is comprised of three subscales: administrative support, professional sharing and support, and professional commitment. No further information is provided. It seems likely that supervisory behaviors would be a factor in administrative support, but it is not clear to what degree.

Finally, Smith (2005) conducted a study to identify the relative importance of organizational, job, and individual characteristics associated with child welfare job retention. Specifically, it was hypothesized that "staff who perceive their organizations and/or supervisors to be supportive and those who find intrinsic value in the work are … more likely to be committed to their jobs and less likely to leave" (p. 158). As noted above, Smith found that supervisory support was significantly positively related to job retention as were extrinsic rewards including the facilitation of work-life balance. However, intrinsic job value was not found to be a significant predictor of retention. Instead, organizational level characteristics including average caseload size and average turnover rate were found to substantially affect the likelihood that a worker would remain employed (or not).

Smith notes that one of the limits to her study is that only 12 county-level units were included in the analysis. Therefore there were limits to the level of power for the multi-level logistic model on Level 2. More study subjects would have strengthened the model. Additionally, all of the study cites had been experiencing relatively high turnover. The lack of variability in the organizational variables may mean that significant predictors of retention were not adequately represented in the sample.

Finally, Smith notes that while the instruments she used had adequate face validity and appropriate levels of internal consistency reliability, they were not standardized. The study was initially designed solely to collect information about turnover, not to test theory-based hypotheses and therefore the use of standardized measures was not a priority initially. For the purposes of this paper, the instrument of most interest is the scale to assess supervision. Smith operationally defined "supportive supervision" through the use of a 24 item scale. Factor analysis suggested the scale consisted to two dimensions: supervisor supportiveness and supervisory competence. Supervisor supportiveness was the mean score of 13 items including "My supervisor genuinely cares about me," "My supervisor shows approval when I succeed," and "My supervisor is available to me when I ask for help." Supervisor competency was the mean score of 11 items including "My supervisor demonstrates consistency in decisions making," "My supervisor demonstrates leadership," and "My supervisor is knowledgeable about ways to work with children and families. Smith concluded that supervisory competence is an important component of "perceived supervisory support" and used the mean score of the entire scale as her variable called "Supervisor support." This is another example of the way in which a finding regarding the importance of supportive supervision may be somewhat misleading. When using Kadushin's terminology, the types of supervisory behaviors included in Smith's scale would include administrative and educational as well as supportive activities.

As is apparent in this brief review of several major studies regarding child welfare worker turnover or retention, findings can be confusing or even contradictory. While multiple theoretical models and research questions may, ultimately, help develop strategies for alleviating the problem of work turnover, the use of standardized measures would significantly increase the comparability of the results. In particular, while multiple studies indicate that "supportive"

supervision plays an important role in retention, inconsistent instrumentation means that it is still not clear exactly which supervisory behaviors are associated with retention or turnover.

Child welfare supervision and quality of services. For the purposes of this paper, it seems important to note that there have been no studies in the field of child welfare that directly link supervisory behaviors with client outcomes. There have been several studies, on the other hand, that looked at issues other than worker retention in an attempt to explore how intervention at the supervisor level might affect quality of services.

In their 2002 study on transfer of training among child welfare workers, Wehrmann et al. found that peer support played a key role in workers use of skills learned at training. They found that supervisory support was an important factor in transfer of training, but not as important as they had hypothesized. They concluded that the supervisor plays an important role in supporting a climate in which peer support was available to workers in order to maximize transfer of training.

Employees tend to report that supervisors rarely ask them about their training experience and fail to demonstrate any expectation that the employees behavior will improve (or change) as a result of the training (Wright, 2003). Additionally, child welfare workers' training needs are often inadequately assessed; workers are forced to attend mandatory training on material with which they are already familiar or which has little apparent relevance to their work (US GOA, 2003; Wehrmann et al., 2002). Because the trainee's previous experience with training plays a role in the individual's motivation to transfer training (Bennett, Lehman & Forst, 1999) these negative training experiences can have longer term negative consequences for training transfer.

As mentioned above, the SR QIC, 4 state supervisory training (learning collaborations) demonstration projects included a program evaluation measure to assess the degree to which

outcomes for children and families were affected. Unfortunately, it was not ultimately possible to say to what degree the project had impacted outcomes; there were too many potential intervening variables. In a paper reviewing the implementation of the demonstration project in Tennessee, Steppe and Jones (2007) note the challenges they encountered included: changes in leadership within the Department; shifts in organizational priorities; participants who may have felt forced to participate because they were selected; problems with the quality of mentors selected (a key element in the intervention model); high profile cases involving children who were seriously injured: and lawsuits (which dictated substantial changes in operating procedures). All of these factors impacted participation levels of supervisors in the program and were likely to have an effect on their ability to transfer the material learned in training into the agency. Other priorities likely impacted effort and attention paid to this endeavor. While a comparison group would to some degree control for effects of these challenges on workers and workers' practice, there was no way to control for the ways in which the intervention was compromised.

Supervision from the Perspective of the Supervisee

Although research related to supervisory practices in child welfare has largely focused on supervisor self-report or single dimension measures of worker perceptions, there are examples of more complex evaluations of supervisory practices and of supervisees' experiences of supervisory practices found in the literature in related practice areas and disciplines. In the field of psychology, clinical supervision plays a key role in the training and development of psychotherapists. In recent years, a growing number of studies have attempted to assess the relationship between psychotherapy supervisors' behaviors and supervisees' satisfaction with supervision, acquisition of clinical skills, and perceptions of self-efficacy.

Ladany, Lehrman-Waterman, Molinaro, and Wolgast (1999) conducted an investigation of 151 psychotherapy supervisees perceptions of their supervisors' adherence to ethical guidelines. They were interested in the effect that perceived violations affected supervisory working alliance and satisfaction with supervision. The researchers constructed a Supervisor Ethical Practices Questionnaire (SEPQ) and a Supervisor Ethical Behaviors Scale (SEBS). They also used the Working Alliance Inventory-Trainee Version (WAI-T) developed by Bahrick in 1990 (as cited by Ladany, Lehrman-Waterman, Molinari & Wolgast, 1999) and the Supervisory Satisfaction Questionnaire (SSQ) previously developed and utilized by Ladany, Hill, Corbett, and Nutt in 1996, and Larsen, Attkinson, Hargreaves and Nguyen in 1979 (as cited by Ladany, Lehrman-Waterman, Molinari & Wolgast, 1999). Fifty-one percent of supervisees reported at least one ethical violation by their supervisors. Findings included that greater non-adherence to ethical guidelines was significantly related to a weaker supervisory alliance and lower satisfaction with supervision.

Lehrman-Waterman and Ladany (2001) investigated the relationship between psychotherapy supervisors' specific evaluation practices and supervisor working alliance and supervisor satisfaction. Findings included that more effective goal-setting and feedback practices were related to stronger working alliances and increased trainee satisfaction.

Focus of Study

The focus of this study is to develop a reliable and valid measure which operationalizes the practice of social work supervision in the public child welfare agency. Further empirical assessment of the relationship between child welfare supervision and a variety of workforce issues is greatly needed. The development of a standardized measurement of child welfare workers' experiences of supervisory behaviors would address a significant problem in the current

literature and its use in future studies would permit comparisons to be made between studies that are impossible currently. Finally, as significant funding is currently being invested in the development of curricula and the training of child welfare supervisors, an instrument for evaluating the impact of these programs is needed.

The political and economic climate in which the child welfare system operates is largely concerned with obtaining the best possible outcomes for the least possible financial investment. It is essential that efforts to improve the quality of services provided to vulnerable children and families through the development of a stable, experienced and highly trained workforce be demonstrable. This study aims to develop a tool which can be used to better evaluate supervisory behaviors and refine efforts directed towards improving supervision as a means of developing such a workforce.

Chapter 3: Methodology

While much has been written across a variety of fields regarding scale development procedures, there continues to be variety in the way in which each of these steps is handled, both in terms of the procedures used by researchers and what they choose to report about them (Hinkin, 1995: Worthington & Whittaker, 2006). There is typically agreement, however, that the process of scale development includes: a) determining what is to be measured b) generating an item pool c) determining the response format of the measurement items d) having the initial item pool reviewed by experts e) considering inclusion of validation items f) administering items to a development sample g) evaluating the items and h) optimizing scale length (DeVellis, 2003; McDonald, 1999; Spector, 1992). The following is a brief review of literature related to the "best practice" methods for scale development. This review informs the study described later in the chapter.

Definition of a Scale

In general, the goal of most quantitative research is to better understand the relationship between specific concepts. Concepts must be defined in measurable terms (operationalized) to become research variables. The more concrete or unambiguous the concept is (e.g., age, marital status), the more straight-forward its measurement is. For items like age a single item will result in an accurate measure of the concept. However, theoretical variables which require a respondent to reconstruct, interpret, judge, compare or evaluate information that is not readily

accessible are ideally suited for the use of a scale in measurement. The complexity of this type of variable could not be captured through the use of a single item; multiple items or questions would be required (DeVellis, 2003; Rubin & Babbie, 2001; Spector, 1992).

In the social and behavioral sciences many concepts of interest are not directly observable and can be complex and ambiguous. Social and behavioral theories guide the researcher's conceptualization of constructs which are believed to represent the concept of interest. Paper-and-pencil measurement scales are widely used to operationalize complex constructs. A summated scale contains multiple items which will combined (summed). Each item has an underlying value which can vary quantitatively (e.g. strongly agree to strongly disagree). Additionally, there are no "right" answers; a summated scale is not used to test knowledge or ability. A scale is a collection of items that are combined into a composite score intended to assess levels of theoretical variables which are not otherwise readily observable (DeVellis, 2003; Spector, 1992).

Scale Versus Index

Scales are distinguished from other types of multi-dimensional measures that yield a composite score. Scales consist of multiple items whose values are caused by a single underlying factor or construct (DeVellis, 2003). For example, the respondent's answers to individual items comprising a job satisfaction scale are related to a common cause, the current fit between the employee's expectations about their job and the employer's ability to meet those expectations. All of the items in a scale contribute to the measurement of one underlying construct. On the other hand, rather than measuring the effect of an underlying single construct, indexes incorporate multiple items to measure several constructs or a multidimensional construct. For example, a measure of a candidate for employment's success might include

attitudes towards work, age, education, training experiences and various aspects of personality. These items are not linked by a common concept, but together form an index or a multidimensional construct: the likelihood of the candidate contributing positively to the organization.

Conceptualization

The first step of scale development is to determine as clearly as possible what it is the researcher wants to measure. The challenge of this step is frequently underestimated, as many researchers think they have a clear idea of what they want to measure, only to discover that the construct is more ambiguous than they thought. Substantive theories related to the phenomena to be studied are helpful in guiding the conceptual process as they provide clarity and specificity (DeVellis, 2003; Spector, 1992). Additionally, the scope of the measurement is very important. A scale can vary along a number of dimensions, including content domains (e.g. depression versus overall psychological adjustment), setting (e.g. questionnaires designed specifically for particular environments), and population (e.g. children versus adults or child welfare workers versus college students). A well defined construct is essential to the development of a good measurement. A poorly defined construct may lead to the inclusion of items that may only be peripherally related or the exclusion of items that address important components of the construct of interest (Worthington & Whittaker, 2006).

Development of an Item Pool

Relationship of items to underlying variable. Once the construct to be measured has been well-defined, the next step is to develop a large pool of items from which the scale components will be drawn (Spector, 1992; Worthington & Whittaker, 2006). Each item should reflect the construct of interest and be considered, individually, to be a test of the strength of the

latent variable. Altogether the items should constitute a more complete and more reliable measure of the construct than each item on its own (DeVellis, 2003; Spector, 1992).

Number of items. When developing an initial item pool, it is generally considered appropriate to be redundant. The ultimate goal is to develop a final scale that is adequately reliable and valid, and to that end, testing various ways of expressing similar ideas in the development process should contribute to better psychometric properties in the final scale. It is not uncommon for draft scales to have 3 to 4 times the number of items anticipated for inclusion in the final version (Worthington & Whittaker, 2006).

The number of items developed should be a reflection of the complexity of the construct to be measured. While it is not possible to make an across-the- board recommendation regarding the number of items that should be included in an item pool, the pool should always contain considerably more items than desired in the final scale. Internal consistency reliability is a function of how strongly the individual items correlate with one another. With a larger pool, items can be retained or deleted in order to end up with a final scale with strong internal consistency. Additionally, in the process of assessing factorial validity, items which do not contribute meaningfully to the factor solution may be deleted. In other words, a large pool is better than a small pool in the item development stage. The exception to this rule noted by DeVellis (2003) is that a pool that is too large to be administered on a single occasion is too large.

In his review of the scale development in the field of organizational development, Hinkin (1995) reported that the process of item generation is often poorly addressed in the literature. He suggests that this step is poorly understood and that, too often, mistakes in the item generation stage result in inadequate measures. He identifies two item generation methods: deductive and

inductive. Deductive scale development is dependent upon a thorough review of the literature and the development of a comprehensive classification schema based upon the theoretical definition of the construct to be measured. The researcher then uses the domains of the schema to develop related items. Alternately, the researcher can utilize a sample of respondents who are experts in the subject matter to provide examples within the domains which can be subsequently used to develop items.

The second method identified by Hinkin does not depend on theory, but rather begins by asking a sample of respondents to describe their feelings or perceptions regarding their organization or some aspect of behavior. The researcher then uses content analysis to begin to develop responses into items and related items are classified into domains. Both methods presume that a sorting process will take place prior to the first administration of a draft scale in order to eliminate items which ultimately are not a good fit with the construct to be measured. However, both methods are considered an important first step towards developing a scale with reasonable content validity.

Characteristics of good and bad items. In general, good items should be unambiguous, concise, and readable (DeVellis, 2003; Spector, 1992). When developing a scale, items which are exceptionally lengthy are generally too complex and lacking in clarity; unnecessary wordiness should be avoided. The reading level of the population of interest should be considered. Generally, a reading level between the fifth and seventh grades is appropriate for most instruments used with the general population. Reading level can be evaluated by assessing the length of the item sentences and the number of syllables in each word; an average sentence at the six grade level has approximately 15 words and a total of 20 syllables (DeVellis, 2003). Multiple

negatives and advanced vocabulary or jargon should also be avoided in item construction (Rubin & Babbie, 2001; Spector, 1992).

Double-barreled items are particularly problematic as they require the respondent to endorse two ideas, rather than one. Each item should contain only one idea. For example, "my supervisor is good at her job because she is experienced" forces the respondent to determine the degree to which he or she agrees that the supervisor is good at her job and *also* the degree to which he or she agrees that the supervisor is experienced. What if the supervisor is very experienced, but the respondent doesn't think she is good at her job? How should the respondent answer? Another problem to be avoided is ambiguous pronoun usage. In the item "employees should not be allowed to join unions because they cause conflict in the work place," do the employees or the unions cause the conflict? Because there is more than one way the respondent can interpret the pronoun, the researcher will not be able to interpret the respondents' answers with certainty.

As a rule, good grammar is an important component in the construction of clear and unambiguous items (DeVellis, 2003; Spector, 1992). Although cautions against poorly worded items abound in the scale development literature, it is not uncommon for double-barreled items or items which do not appear to be conceptually consistent to appear in scales in published articles. These items often lower the scales reliability estimates and fail to load as expected in factor analysis and, if this level of analysis is conducted, are generally eliminated in the item analysis stage (Hinkin, 1995).

Most scales include both positively and negatively worded items; that is, items which assess for the presence of the construct and items which assess its absence. Including both types of items is intended to avoid acquiescence or agreement bias, defined as the respondent's

tendency to agree with items, regardless of their content. Acquiescence bias would be observable in instances where a respondent indicates high levels of agreement with both positively and negatively worded items. However, including both positively and negatively worded items can be confusing to respondents and may, in fact, contribute to individual items performing poorly (Hinkin, 1995). Negation in items is easily overlooked, and may further contribute to poor results (Rubin & Babbie, 2001); "my supervisor is not good at her job" may be read as "my supervisor is good at her job". DeVellis (2003) recommends against using items worded in the opposite direction as he has found the disadvantages to outweigh potential benefits.

Response format for measurement. The response format of the final scale must be considered in tandem with item development as the items in the item pool must be consistent with the chosen format. Does the format require the development of declarative statements or lists of positive and negative characteristics associated with the construct? In general, formats which allow the scoring of the items on some continuum and that can be summed to form a scale score are the most compatible with statistically based methods of evaluating data (DeVellis, 2003; Spector, 1992). This review of the steps for scale development is focused on rigorously tested, theoretically based instruments, and therefore response formats reviewed will be those most compatible with higher level statistical analysis and acceptable quantitative research standards.

Types of scales. In Thurstone scaling, items are developed which correspond to different intensities of the underlying latent variable. Respondents indicate their agreement with agree-disagree options. Ideally, the items are spaced to represent equal intervals between levels of intensity, so that mathematical procedures based on interval scaling could be used (Spector,

1992). Unfortunately, developing items which correspond to equally spaced levels of a particular attribute or construct is not an easy process; practical problems tend to outweigh the advantages (DeVellis, 2003).

A second type of scale format that relies on ordered items is Guttman scaling. A Guttman scale is a series of items which correspond to progressively higher levels of an attribute. Responses are agree-disagree, with an affirmative response indicating agreement with the individual item and all the items corresponding to lower levels of intensity. With a Guttman scale, it is the point of transition from agree to disagree or disagree to agree that is the focus of data collection. The development of a Guttman scale can be labor intensive and expensive, and, as result is not widely used. Additionally, it is not compatible with the most widely used measurement theories (DeVellis, 2003). In both the Thurstone and the Guttman type of scale, the assumption of equally strong causal relationships between the construct to be measured and each of the items does not apply (Spector, 1992).

The most common type of scale and the one most appropriate for use with the statistical measures referenced later in this review is a scale with equally weighted items. Each item is an indicator for the underlying latent variable and can be combined by simple summation into an acceptably reliable scale (DeVellis, 2003).

The semantic differential scaling method is chiefly used to evaluate participants' responses to one or more stimuli. For example, the stimulus might be a group of people, in this case, politicians. The target stimulus is followed by a list of paired adjectives with seven or nine blank lines between them, each representing the opposite ends of a continuum: "honest… dishonest," : "fair…unfair," and so on. The respondent marks the blank which corresponds to his or her belief about the target stimuli relative to each continuum (DeVellis, 2003; Rubin &

Babbie, 2001). The visual analog method is similar to the semantic differential method, except that a continuous line is used between the two end points. The respondent marks a point on the line to correspond with their response. A major advantage of an analog scale is that it is very sensitive. However, a disadvantage is the problem of how to assign values to responses in an analog scale.

Another common response format is the binary option. In this type of scale the respondent may be asked to check agree or disagree in response to items. Alternately, the respondent may be provided with a list of elements and asked to check those that apply. A major disadvantage of this method is that there is very little variability on each item. On the other hand, each item is easy for the respondent to complete and so respondents may be more willing to complete more binary items than they might items to complete items which require greater concentration and discrimination (DeVellis, 2003).

The mostly widely used scale format is the Likert scale. In a Likert scale, the item is presented as a declarative statement followed by response choices which indicate varying degrees of agreement with the statement. An even number or an odd number of response options can be provided. Response options should be worded so that there are roughly equal intervals between options in respect to agreement with the statement. For example, a six response option set would include: “strongly agree,” “moderately agree,” “Mildly agree,” “mildly disagree,” “moderately disagree,” and “strongly disagree.” A midpoint of either “neither agree nor disagree’ or “agree and disagree equally” can also be used.

Likert scaling is widely used for instruments measuring opinions, beliefs or attitudes. The goal of these measurements is to distinguish between people with strong or moderate positive and negative opinions. In order to be effective, the statements used in a Likert scale

need to be clearly stated assertions of belief or attitude which are not overly mild or overly extreme. The response set choices should provide the opportunity for the complete range of attitudinal of responses to the statement (DeVellis, 2003; Spector, 1992).

Response format time frame. Finally, in developing a response format, the researcher must decide on the time frame for the items. This can be implied or explicit; each will have bearing on the item stems and the directions provided the respondent. Depending on the construct being researched, it may be more appropriate to have a universal or global time frame. In these instances the assumption is that the respondents' beliefs will be enduring or stable over time. In other cases, measures are designed to assess relatively transient phenomena (DeVellis, 2003). The researcher may specifically be interested in attitudes, behaviors or beliefs present in a specific time period: "the last three weeks," "the past week," or in the period since a particular event. The response format time frame should be consistent with the theoretical basis which underlies the development of the individual items.

Number of response options. Scale items generally consist of two parts: a stem and a series of response options. Response options may include using a number from 1 to 100 to provide for numerous discrete responses. Alternatively, response options may be limited to a few choices: agree and disagree; or none, a little, a moderate amount, and a lot, for example. In evaluating response options, it is important to remember that an important aspect of a scale is its variability. If a scale cannot discriminate between differences in an underlying attribute, its utility is limited. Two strategies for insuring opportunities for variability include 1) utilizing many items and 2) providing many response choices within items.

However, it is also important to balance opportunities for variability with ease of completing the scale. Too many items or too many response choices may confuse or discourage

a respondent. Additionally, in order for respondent's choice to be meaningful, there should be real and meaningful differences between the response choices. When asked to rate something on a scale of 1 to 50, for example, respondents may choose to respond in multiples of 5, effectively reducing the number of options. Further, differences between a response of 35 and 37 may not reflect an actual difference in the attribute being measured. On the other hand, asking a respondent to discriminate between vague response options (e.g. very few, few, some, many, etc) may also create problems. It may be preferable to have fewer response options than to use ones that are ambiguous.

The use of mid-point in response options should be decided in relation to the construct one is attempting to measure. A mid-point or neutral response may foster equivocation and for some constructs this may not be desirable. For example, when asking respondents to evaluate themselves in comparison to others, the choice of "equal" may permit equivocation. On the other hand, a mid-point may be crucial in determining when two aspects of an attribute are evenly balanced. An example of this would be when a "no preference" option is utilized (DeVellis, 2003).

Likert scales can have as few as 3 response options and as many as 11. Given the wide spread use of Likert scales in both behavioral research and marketing research, there has been considerable interest in evaluating the benefit of a particular number of response options (Bardo, Yeager & Klingsporn, 1982; Dawes, 2008).

Oaster (1989) found that per choice point from 3 to 7, there was a related increase in alternate-forms estimates of stability. His conclusion was that an instrument's reliability could be improved through increasing the number of alternatives per choice point. On the other hand, Bardo, Yeager and Klingsporin (1982) administered a questionnaire of contentless items with a

variety of response formats to a large group of undergraduate students, instructing them to produce random answers. Analysis of the results indicated that for 4, 5, and 7 point Likert scales, the central tendency errors tended to increase and the variances expected tended to be reduced as the number of response options increased. They ultimately recommend use of a 4 point Likert scales due to additional analysis which suggests that the 4 point response option is less susceptible to individual systemic bias which produces an inflated reliability assessment. Bardo and Yeager (1982) used the same study data to look at response style as determined by the correlation of the means of each response format group. They found that Likert-type formats were similarly affected by response style regardless of number of categories in the format. They conclude that increasing the number of categories might increase systemic error, thereby reducing variance rather than increasing it.

Dawes (2008) when comparing 5 point, 7 point and 10 point Likert response sets found that there was no significant difference in variation about the mean, skewness or kurtosis between the three formats and therefore all formats are comparable for use with statistical tests including regression, confirmatory factor analysis, structural equation models. However, use of the 10 point format resulted in a significantly lower mean when compared to the 5 and 7 point formats. He concludes that a scale with more response options produces "slightly lowered scores," relative to the upper limit of the scale, which should be taken into consideration when comparing results using differing response formats.

Similarly, Rhodes, Matheson and Mark (2010) when comparing results from the randomly assigned use of 5 point Likert, 7 point Likert, 5 point semantic differential scaling and 7 point differential scaling in a study of attitudes towards physical activity and actual behaviors, found that there was increased variability in the data with both 7 point formats. They further

found that the 7 point Likert format had the highest overall reliability. However, they did not find that the differences in format translated into any meaningful difference in predictive validity of actual behavior. They conclude that all 4 formats are appropriate for use because of their relatively equivalent outcomes.

Expert Panel Review

Once the response format has been decided upon and an initial pool of items has been developed, the next step of scale development is an initial review of the items by a panel of experts. The experts who comprise the panel should be individuals who have some expertise relative to the construct of interest. This process serves several functions: to address basic grammatical and editing "clean-up" of items; to review the draft scale's face validity (Hardesty & Bearden, 2003); and to do an initial assessment of the scale's content validity (DeVellis, 2003; Worthington & Whittaker, 2006).

The expert panel plays an important role in reviewing the items for grammatical errors and lack of clarity. The panel can help identify alternative readings of items that are ambiguous or confusing, permitting the researcher to eliminate them or rewrite them as appropriate. Redundancy and readability can also be addressed. The panelists may make suggestions regarding the administration instructions, length, and organization of the proposed draft scale (Worthington & Whittaker, 2006).

In Hardesty and Bearden's (2003) review of the use of expert panels in marketing research, they note that how expert panels are used to assess face validity is rarely reported upon. As a result, little is known about the optimal number of expert panelists. In their review of approaches for estimating content validity in nursing research, Wynd, Schmidt and Schaefer (2003), similarly, do not recommend a specific number of panelists or a formula for determining

an appropriate number, although they do comment on the relationship between the number of experts and the statistics used to estimate content validity; more raters increases the risk of chance agreement. In their study, they used 8 experts. DeVellis (2003) also does not recommend a specific number or formula for determining how many experts to use.

Face validity. Face validity refers to the researcher's assessment that each item is an accurate reflection of the construct to be measured. In their review of the use of expert judges to improve face validity of scales developed for consumer and marketing research, Hardesty and Bearden (2003) liken face validity to a dart board. Each item should "hit the dartboard." While face validity is not considered to be a strong measure of the scale's overall validity, the researcher's goal is to start with strong face validity. Prior to using an expert panel, each item should be scrutinized and carefully assessed for face validity by the researcher; however, the expert panel can provide additional information toward assessing face validity. It is possible that while the researcher's intended meaning for an item is a good fit with the underlying variable, the reviewer's reading of the item yields a different meaning (DeVellis, 2003).

Hardesty and Bearden (2003) recommend that the expert panel be asked to review each item in terms of its "representativeness" of the construct as defined by the researcher. They suggest that the "sumscore" decision rule be utilized to determine inclusion or deletion of items in support of greater face validity. The sumscore decision rule is based upon Zaichowsky's (as cited in Hardesty & Bearden, 2003) rating structure which requires that each judge rate each item as "completely," "somewhat," or "not at all representative" of the construct to be measured. For each judge that assigns "completely representative" the item receives a score of 3 points. For an assignment of "somewhat representative" the item receives 2 points. If assigned "not representative" the item receives 1 point. The total item score is then summed and averaged. If

the item averages a "somewhat representative" score of 2, then the item is retained. Hardesty and Bearden compared this method with several others and found it to be more predictive of items which are eventually contained in the final scale than two other uses of an expert panel in assessing face validity.

Content validity. Finally, an expert panel can facilitate the assessment of content validity. Content validity is the extent to which a specific set of items reflects the construct domain. Hardesty and Bearden (2003) distinguish content validity from face validity as the degree to which "darts… land randomly all over the board to obtain a proper representation of the construct" (p. 99). The goal of the researcher is to include items which adequately define the construct in question and nothing else. Ideally, the items should include all aspects of the latent variable while avoiding the inclusion of any items not related to the construct in question. Content validity is easiest to assess when the construct to be measured is well defined; it is more difficult to assess when measuring beliefs, attitudes or perceptions. The expert panel can be helpful in noting content areas which may have been overlooked in item development or when items may stray from the formal definition of the construct to be measured (DeVellis, 2003).

In their review of approaches for estimating content validity in nursing research, Wynd et al. (2003) note that the specific use of an expert panel is often minimally addressed in publication. They present two quantitative approaches for improving the transparency and empirical quality of the expert review process. The first is the CVI, which evaluates the proportion of agreement between raters. A 4 point Likert scale is used to rate the items as 1= not relevant, 2 = somewhat relevant, 3 = quite relevant and 4 = very relevant in terms of how well the item describes the construct in question. Scores are then compressed into two dichotomous categories, content invalid and content valid. The researcher next tallies the proportion of cases

in which raters agree. The reported CVI is the proportion of items that are considered content valid (Lynn as cited in Wynd et al., 2003). According to Wynd et al., this approach is promoted by nurse researchers although it has been criticized due to its potential for overestimating agreement; CVI lacks a "no agreement" category and because only two categories are compared, random chance could affect the level of agreement. Additionally, the use of more experts can increase chance agreement.

The second method discussed is the use of the multirater Kappa statistic as an index of interrater agreement. Cohen's (as cited in Wynd et al., 2003) coefficient kappa removes chance agreement from the proportion of agreement. P_e is the proportion of agreements expected to occur by chance alone, and $(P_o - P_e)$ represents the instances of "real" agreement.

$$k = \frac{P_o - P_e}{1 - P_e}$$

Although this use is an extension of kappa beyond Cohen's original intentions, it is mainly used to test interrater agreement when observers are choosing between dichotomous categories. The values of kappa range from +1.0 to -1.0 with a +1.0 indicating perfect agreement, 0 indicating agreements are equal to what could be expected by chance and a negative kappa indicating that the raters disagree more often than what might be expected due to chance. Wynd et al. conclude that the use of kappa offers additional information beyond use of the CVI and recommend the use of both in order to increase confidence in the content validity of new instruments. However, since both procedures they discuss address the relevance of an item to the construct for which a scale is being constructed, it is not clear how either procedure offers more than a formalized means to assess face validity; neither procedure addresses the degree to which the items comprehensively cover the scope or domain of the construct.

When a proposed scale consists of several well defined domains or subscales, a common method of assessing content validity is to have reviewers classify the items into the domains (Hardesty & Bearden, 2003; Hinkin, 1995). Often the category of "other" is also included. Items which are classified correctly at least 80% of the time are retained. The experts' ability to determine to which category the items belong is a strong indicator of content validity (Hinkin, 1995).

In any case, both face and content validity are only preliminary assessments of an instruments' ultimate utility. Procedures to assess face validity insure that the items *appear* to be consistent with the construct to be measured. However, when developing a new scale for a construct which has not been previously measured, there is understandably some question as to what items measuring that construct look like. Face validity alone is not an indicator of validity. Additionally, there is no agreed upon criterion which definitively determines whether an item list has content validity. While it is desirable to have content validity, there is no meaningful way to assess whether all of the items necessary to measure the full domain of a construct have been included. Content validity cannot be rigorously assessed (Carmines & Zeller, 1979).

Ultimately, it is the researcher's responsibility to determine which items stay in the scale and which are eliminated. Expert panel review can be helpful in providing the researcher with additional information towards making these decisions. Indeed, it is not uncommon to use more than one phase of judging and to employ more than one decision rule (Hardesty & Bearden, 2003). Items that are consistently found to have low relevance can be removed; items which are consistently found to be relevant are presumed to increase the scale's validity.

Inclusion of Validation Items

The scale developer's goal, ultimately, is to end up with a list of items which have been assessed as having reasonable validity and reliability to constitute the scale's use as a "good" measure of the phenomenon being studied. There are two types of items which the researcher may decide to include in the same questionnaire which can be helpful in evaluating the validity of the scale under development.

The first of these are items which serve to detect problems in the scale. Of particular concern would be the problem of respondents' answers reflecting considerations beyond those the scale is attempting to measure. One such consideration is social desirability, or the motivation of the respondent to be viewed positively by the researcher. There are social desirability scales which can be inserted into the developing scale to facilitate the researcher's evaluation of how strongly individual items are influenced by the respondents' motivation to be perceived positively (DeVellis, 2003; Rubin & Babbie, 2001). Individual items which are strongly correlated with the social desirability score obtained could be eliminated from the final measure.

The second type of items to be considered for inclusion at this stage are those which when administered would yield information about the construct validity of the measurement. Construct validity is the assessment of how well the scale performs in relationship to other measurements using hypotheses generated by the underlying theoretical model (Carmines & Zeller, 1979). For example, if a scale is designed to measure depression, scale scores should be highly positively correlated with other established measures of depression. Similarly, the scale scores should not be correlated with factors which are unrelated to depression. Construct validity can be assessed through a separate validation study after the final version of the scale is

established. However, it may be possible to include validation items at this stage. DeVellis (2003) suggests that if the sample is sufficiently large, it be split in half, or into larger and smaller subsamples so that the secondary subsample can be used for analysis of construct validity.

Worthington and Whittaker (2006) caution against including additional scales with the first administration of a draft scale. Their concerns include instrument length and respondent contamination. As noted previously, scale development studies may include 3 to 4 times the number of items which will eventually comprise the scale. Adding additional questions to a questionnaire which may already be quite lengthy is likely to have a negative effect on the completeness of the results. Secondly, the items from existing measures might influence the participants' responses to items for the new scale. It would be very challenging, if not impossible, to control for the effects of different measures while also testing the factor structure and initial reliability estimates of the new scale. They recommend that assessing convergent and divergent validity occur in a later stage of scale development.

Factor analysis is another method which is used to evaluate the relationship between the items in the new scale and the theoretical model describing the construct to be measured upon which the items are based. Factor analysis allows the researcher to determine how many latent variables underlie a set of items and the substantive content or nature of these latent variables (DeVellis, 2003; McDonald, 1999; Spector, 1992). This process will be described in more detail below, but is of note here because factor analysis does not require the inclusion of additional items during the initial administration of the new scale although the factor solution can suggest good fit or poor fit with the theoretical model (as does construct validity). In fact, given sufficient sample size, both exploratory (EFA) and confirmatory factor analysis (CFA) utilizing

structural equation modeling (SEM) can be conducted to further assess the validity of the new scale (DeVellis, 2003; Grimm & Yarnold, 1995; Worthington & Whittaker, 2006). The CFA process will not be discussed much further as it is beyond the scope of the study presented here.

Finally, another approach to evaluating validity is to use demographic information that is collected to describe the sample to determine the degree to which demographic variables which are theoretically related to the underlying construct are correlated or fail to correlate with the proposed scale. In essence, using multiple regression, a model composed of predictor variables can be tested to see the degree to which these variables explain the variance in the scale scores. Theoretically based hypotheses would be proved or not, suggesting that the scale scores are consistent with the theoretical model. This approach has the advantage of not requiring the inclusion of additional scales. However, the results would clearly be only a first attempt to assess validity. Certainly estimating the new scale's validity is an important aspect of scale development. The pattern of relationships which results from an initial administration of a new scale will support claims of validity or indicate potential problems or areas to be further assessed in later stages of scale assessment.

Administration of Items to a Development Sample

Once the researcher has determined which scale items and/or validation items will be included, the next step is to develop a questionnaire and administer the new scale to a pool of respondents.

Sampling method. In their review of scale development research in counseling psychology, Worthington and Whittaker (2006), address methods for selecting a developmental sample. They note that representativeness is not considered to be crucial as long as the sample includes those who would score low and score high in the target population and the sample is

sufficiently large. However, they note the importance of evaluating the sample relative to specific the populations of interest to determine the degree to which certain segments of the population of interest may be missing. They found that the most common form of sample generation used was purposive sampling from a specific target population. Other methods used included a combination of convenience and purposive and convenience alone. Each of these methods was determined to be satisfactory, provided the development sample generated was large enough.

Sample size. Generally, for scale development, a large sample is recommended. The goal is to have a large enough sample that certain segments of the population are not excluded. Of concern is that a small sample may differ qualitatively from the target population; a particular attribute or opinion relevant to the scale's content may be present or absent in a small sample when it is absent or present in the target population (DeVellis, 2003). Additionally, the random effects of larger samples tend to cancel the scale variance attributable to specific participants. Larger sample sizes are more likely to result in more stable relationships among the variables and, therefore greater replicability in validity estimates (Grimm & Yarnold, 1995; Worthington & Whittaker, 2006).

As mentioned previously, if the sample size if sufficiently large, the questionnaire can be administered once, but the sample then divided into two groups, in order to first evaluate and refine the scale and then, using the second sample and the refined version of the scale, perform additional analysis of validity (DeVellis, 2003). This might include a confirmatory factor analysis as well as an analysis of the pattern of relationships between the scale and other constructs. However, having a large enough sample size is a significant barrier to completing this level of evaluation in one study.

Worthington and Whittaker (2006) provide four principals for determining the appropriate size of the development sample: 1) a sample size of 300 is generally adequate, 2) sample sizes of 150 to 200 are likely to be adequate if upon factor analysis, the data set yields communalities higher than .50 or 10:1 items per factor with factor loadings of 0.4, 3) smaller samples may be adequate if factor analysis yields communalities of 0.6 or greater and or at least 4:1 items per factor and factor loadings greater than 0.6, and 4) a sample size of 100 or fewer than 3:1 participant to item ratio is insufficient. Unfortunately, these criteria can only be assessed after the data has been analyzed. They are suggested as indicators that the sample is inadequate and that additional data should be collected. Hinkin (1995) suggests that a sample of 150 is adequate for exploratory factor analysis, while 200 is preferable for confirmatory factor analysis.

DeVellis (2003) recommends at least 300 subjects be used. He qualifies this by noting "if only a single scale is to be extracted from a pool of about 20 items, fewer than 300 subjects might suffice" (p. 88). However, scale development should generally include factor analysis (discussed below). DeVellis emphasizes that larger numbers of items to be factored and greater numbers of anticipated factors require larger sample sizes. He suggests the use of the Tinsley and Tinsley (as cited in DeVellis, 2003) ratio of 5 to 10 subjects per item is sufficient up to a sample size of about 300, where the ratio can be relaxed.

Evaluation of the Items

Once the new scale had been administered, its properties must be carefully assessed. The data generated will facilitate the researcher's evaluation of the scale's reliability and validity and its appropriateness for use to measure the construct of interest. Additionally, the data can guide the process of refining the scale for future use.

Reliability estimation. The first quality which is assessed is the new scale's reliability. Reliability is essentially the degree to which the respondents will respond to the items in the same way if the measure is repeated. While random error is expected to contribute to some of the variance in the responses, a reliable instrument consistently produces results which are (mostly) free of random error and therefore, considered more representative of the values of the underlying construct (Carmines & Zeller, 1979). Reliability can be assessed in three ways: test-retest, equivalent form and internal consistency (Hinkin, 1995; Rubin & Babbie, 2001)

Test – retest. In test-retest reliability evaluation, the scale being tested is administered once and then administered to the same sample again. The scale's reliability is assessed by the degree to which the responses in both administrations agree. Beyond the issues associated with attempting to retain the members of the sample over time, there are several problems with this method which may result in an inaccurate assessment of reliability. The first is that the respondent may have changed in some way from the first administration to the second. For many scales the ability to capture more subtle differences is desirable. However, if the responses from two administrations of the same scale are substantially different, there may be a problem with reliability or the scale may simply be capturing changes in the respondents (Carmines & Zeller, 1979; DeVellis, 2003; Rubin & Babbie, 2001). This method of reliability estimation is typically only recommended for use with scales which capture data which is considered to be relatively stable over time (Hinkin, 1995). The second problem is that having participated in the administration the first time, the respondent may be induced by that experience to respond differently the second time (Carmines & Zeller, 1979; DeVellis, 2003).

Equivalent forms. The equivalent forms method uses two version of the same scale administered to two different samples and then correlates the scores between the two groups. If

the two versions of the scale are indeed equivalent, and the scores of each group are highly correlated, the scale is presumed to be reliable. However, developing a single scale is challenging enough. The process of developing equivalent scales is burdensome and costly (Carmines & Zeller, 1979). An alternative to equivalent forms is the use of split-half reliability. In a scale of sufficient length, the items can be divided into two groups (usually odd numbers and even numbers) and correlated to assess reliability. If, however, the split halves are not in fact equivalent, then the assessment of reliability may be distorted (DeVellis, 2003; Rubin & Babbie, 2001).

Internal consistency. The final method for assessing reliability is to evaluate the internal consistency of a measure. The researcher's goal is to develop a set of items which are highly inter-correlated as this is an indication that, in sum, they relate well to the underlying construct (Rubin & Babbie, 2001). The assumption is that the higher the individual item reliabilities, the more intimately they are related to the true score (DeVellis, 2003). The more reliable the individual items are, the more reliable the scale that they comprise (assuming that they share a common latent variable).

Item variance. A valuable quality in a scale is a relatively high variance. An item which is answered in exactly the same way by all respondents will not tell the researcher very much about the diversity of the respondents' view relative to the construct being studied; item deletion is likely to be appropriate. Additionally, univariate item analysis permits assessment of potential problems related to missing data. An item that is repeatedly left blank during an initial administration is likely ambiguous or confusing to participants. It may be appropriate to delete "bad" items with a significant number of missing responses prior to conducting additional analysis.

Potential problems can be identified through the analysis of the mean of each items. On a Likert type scale, each item's mean should be located close to the center of the range of values. If the mean falls near either end of the range, it may vary only over a narrow range and will have low variances and the data would be positively or negatively skewed. These items will tend to have lower correlations with other items and will reduce the overall assessed internal consistency of the scale (DeVellis, 2003). Items with low variance are of concern because they likely fail to measure certain values of the construct. There are no guidelines specifically about eliminating items due to low individual item variance, however, awareness of potential issues with univariate data may be useful in interpreting potential problems with multivariate analysis.

Cronbach's alpha. The reliability coefficient, alpha, is one of the most important indicators of a scale's quality. All of the problems with individual items- poor variability, non-central mean, low item-scale correlations and weak inter-item correlations- will contribute to a reduced alpha. Once the weakest individual items have been removed, the alpha is a means for evaluating the items which are left. The alpha most frequently used is the Cronbach's alpha (Cronbach, 1951 as cited in Spector, 1992) which is readily available in SPSS and SAS. The alpha can produce values from 0.0 to 1.0 with 0.0 indicating no correlation and 1.0 indicating high correlation. In evaluating the internal consistency of a scale, below .60 is unacceptable, between .70 and .80 is respectable and above .80 is very good (Bland & Altman, 1997; DeVellis, 2003; Hinkin, 1995; Spector, 1992). While the level for Cronbach's alpha considered acceptable is stated frequently in the literature, studies utilizing untested instruments frequent report alphas in the unacceptable range (Hinkin, 1995). Further, in the developmental stages it is advisable that a higher alpha be sought as items are still being selected or rejected based on their contribution to the overall alpha coefficient. It is possible that some of the covariation among

items may be due to chance. In further use, the alpha may deteriorate. However, if it is a very good alpha to begin with, even with deterioration it should stay in the acceptable alpha range (DeVellis, 2003).

Validity estimation. Although the intention of the researcher is to develop a scale which is strongly associated with the underlying concept he or she is attempting to measure and all of the procedures described above improve the likelihood that the scale will function as intended, the list of items developed may not be inclusive enough or may include items which measure related but different constructs. It is essential at this point to examine the underlying latent variables for the item set. The best means to determine the number of components or factors which account for the variance in the data collected is utilization of factor analysis (Carmines & Zeller, 1979; DeVellis, 2003; Grimm & Yarnold, 1995; Mertler & Vannatta, 2005; Spector, 1992).

Factor analysis. Factor analysis serves three purposes: to help the researcher determine how many latent variables underlie a set of items; to provide a means of condensing information so that the variation can be expressed through a smaller number of variables; and to facilitate the definition of underlying factors. Generally speaking, factor analysis is a procedure used to determine the extent to which shared variance exists among a set of variables. It is a process which determines which variables "hang together;" factors are the groupings of variables which measure a common idea or construct (DeVellis, 2003; Mertler & Vannatta, 2005).

Factor analysis works by assessing how much of the association between variables a single construct (factor) can explain. If it appears that a single factor solution provides an inadequate explanation for the variance among items, a second factor which explains the remaining variance is identified. This continues until only an acceptable amount of variance is

not accounted for. The factor analysis extraction procedure involves the development of an ideal one-factor model for which the inter-item correlations are computed. The actual correlations are then compared to the model-driven correlations. If the correlations are substantially different, the single factor model is found to be inadequate. The next step is then to extract a second factor and to again assess the degree to which there is remaining variance which is not accounted for by either factor. Factors will continue to be extracted until only very small residual correlations are present (DeVellis, 2003; Grimm & Yarnold, 1995).

Principle components vs. common factors. The term factor analysis is often used to refer to both principal components analysis (PCA) and common factor analysis (FA). The PCA extraction process is an analysis of all variance in the data; unique, shared and error variability are all analyzed. In FA, only shared variability is analyzed. PCA analyzes variance. FA analyzes covariance. Additionally, with PCA it is assumed that a model will explain 100% of the variance in the variables, while with FA it is assumed that less variance is explained (Grimm & Yarnold, 1995). Frequently the procedures are used interchangeably and when the scale items have something meaningful in common, use of either procedure has been found to produce similar results (DeVellis, 2003; Worthington & Worth, 2006; Mertler & Vannetta, 2005).

Mertler and Vannatta (2005) suggest use of the PCA is more appropriate for use in scale development where the focus of analysis is exploratory; the researcher is searching for underlying structure. PCA reduces the number of items while retaining as much variance as possible. Its goal is to extract the maximum variance from a data set, resulting in a few orthogonal (uncorrelated) factors. However, Worthington and Whittaker (2006) recommend use of FA for scale development primarily because the outcomes are considered to be more effectively evaluated in a confirmatory factor analysis procedure.

The final test of the quality of the scale will be the replication of the results obtained in the initial administration. Scale development begins with exploratory factor analysis (either PCA or FA). Confirmatory factor analysis is most typically based on structural equation modeling (SEM) which is an extremely useful and flexible method for evaluating how well new data fits the original model (DeVellis, 2003; Worthington & Whittaker, 2006). SEM allows the researcher to specify the relationships among the variables as determined in the previous research and the analysis determines if the data can be fit to that model. A key criterion in determining a scale's validity is the replication of the hypothesized factor structure using a new sample. Confirmatory factor analysis can be used to do this.

DeVellis (2003) suggests that another option for replicating the initial factor analysis results is to conduct a second factor analysis on a new data set. Rather than predict the outcome and test that hypothesis, using exploratory factor analysis again and producing exact or similar results in a new data set is a strong indicator that the factor structure is not the result of error. In both types of replication, however, exploratory factory analysis is the preferred first step.

In any case, this study was a first attempt at developing a scale with appropriate psychometric properties. Confirmatory factor analysis was not planned or conducted.

Extraction of factors. Factor analysis (PCA or FA) consists of two procedures: factor extraction and factor rotation. Factor analysis will extract as many factors as items in the effort to account for all the covariation. However, after a certain point, the factors become less meaningful. In extracting factors, there are several guidelines which are frequently used, the eigenvalue rule and the scree test. DeVellis (2003) states: "an eigenvalue represents the amount of information captured by a factor" (p. 114). If each individual item is a distinct factor, then each would have an eigenvalue of 1.0. Because the goals of factor analysis include reducing the

available information into a smaller number of variables and, in scale development, to determine how many latent variables underlie the items, eigenvalues of 1.0 across the board are not at all desirable. In general, the researcher is looking for factors which capture more information than a single individual item. Generally, it is recommended that only factors with eigenvalues greater than 1.0 be retained.

The scree test is another way of looking at the same issue. In the scree test the eigenvalues are plotted on a graph so that their relative values rather than their absolute values can be assessed. Typically the graph will show a succession of factors with eigenvalues which account for a greater amount of information proportionally than the others and then a series of factors with eigenvalues which are very low and become successively lower. An "elbow" marks the transition from those factors which capture a substantial amount of information to those which fail to. The scree plot rule suggests retaining the factors above the "elbow" (DeVellis, 2003; Grimm & Yarnold, 1995; Mertler & Vannatta, 2005; Worthington & Whittaker, 2006).

Another means of evaluating the factor extraction is to assess how closely the resulting model approaches approximate simple structure. Simple structure is defined as a factor pattern in which several items load strongly on only one factor and the items have zero correlation to other factors in the solution. In exploratory factor analysis, the goal is to approximate simple structure as closely as possible. If factors share items that load highly on more than one factor, these items contribute to a complex structure. Item or factor deletion or both can be used to more closely approximate simple structure (DeVellis, 2003; Worthington & Whittaker, 2006).

Additionally, the larger the number of items that load on a factor, the more likely the factor will be stable in future analysis. Therefore, Worthington and Whittaker recommend deleting factors with fewer than three items, unless those items are highly correlated ($r > .70$) and

relatively uncorrelated with other factors. Because SEM approaches to confirmatory factor analysis assume simple structure, the more closely the exploratory factor analysis solution approaches simple structure the better the likelihood that the solution will be replicated through confirmatory factor analysis.

Mertler and Vannatta (1995) recommend that four criterion be used to determine which factors to retain: a) Eigenvalues b) scree plot c) as many factors as necessary to explain 70% of the total variability and d) assessment of model fit. They further suggest that use of the threshold of an eigenvalue of 1.0 is reliable when the number of variables is >30 and the communalities are >.70 or the number of cases is at least 250 and the average of the communalities is .60 or more.

Extraction methods. Factors can be interpreted in terms of what constructs are associated with them through the analysis of the clustering of items which are "loaded onto" each factor. However, in order to accomplish this, it is advisable to transform the data into a presentation which is easier to understand. DeVellis (2003) likens this to changing the "vantage point" so that the data is laid out in a way that is more immediately understandable. For every set of data there is a hypothetical simple structure where each item loads perfectly on only one factor and not at all on any other. The rotation procedures seek factors that result in each item correlating with only one factor resulting in an organization of the data that reveals natural groupings of items that have something in common.

There are two choices for rotating the data: orthogonal and oblique rotations. Orthogonal rotation assumes that the factors are statistically independent of each other. In this case, the sum of the effects of each of the factors' separate effects equals their combined effects (DeVellis, 2003). This is an ideal factor structure. However, if there is correlation between the factors, then an orthogonal rotation tends to overestimate the factor loadings and may result in the

inappropriate retention or rejection of some items. Additionally, the factor structure may be more difficult to replicate in confirmatory factor analysis (Worthington & Whittaker, 2006).

When the factors are hypothesized or known to be correlated, it is more appropriate to use oblique rotation. In scale development, theory may suggest that the factors are correlated or uncorrelated and researchers frequently use theory to justify their use of orthogonal or oblique rotations. However, it is considered appropriate to evaluate the correlations between factors, particularly in scale development and use the analysis to determine which rotation to use (DeVellis, 2003; Worthington & Whittaker, 2006). If an oblique rotation is specified but the resulting correlations are quite small or less than .15, an orthogonal rotation can be used (DeVellis, 2003).

Interpreting factors. In a best case scenario, the number of factors extracted will match the theoretical model from which the items were derived. In order to determine the nature of the latent factors, it is necessary to examine the items which have the strongest loadings for that factor. When the items are clearly related to each other it is fairly easy to determine the nature of the latent factor. However, when a factor accounts for little of the variance present, items which are seemingly unrelated load equally on a single factor or there is strong cross-loading (items have high factor scores on more than one factor), interpretation is more challenging (DeVellis, 2003; Mertler & Vannatta, 2005). It may be necessary to revisit the item generation phase of scale development in order to produce items which will load appropriately on factors which better fit the theoretical model (Worthington & Whittaker, 2006).

Optimization of Scale Length

The final step in scale development is to refine the scale based on the analysis of data obtained through the initial administration. As noted previously, an initial version of a scale may

have as many as 3 to 4 times as many items as the final version of the scale is intended to have. How many items the final scale should have, though, is not always clear. Hinkin (1995) emphasizes the importance of including items which address all domains or facets of the construct of interest. However, he states that "scales comprised of five or six items that utilize five to seven point Likert scales … would be adequate for most measures" (p. 974).

The goal is to produce a final scale that is optimally balanced between being conveniently short and having acceptable reliability and validity. A longer scale composed of many items with high inter-item correlation which contributes to the overall internal consistency will tend to be more reliable. Additionally, for scales with more items, the inter-item correlation scores can be lower and still contribute to an acceptable alpha value. On the other hand, if a scale is burdensomely long, respondents may fail to complete it. When eliminating items, items which contribute least to alpha can be eliminated first (DeVellis, 2003).

In terms of validity, the scale developer is seeking a parsimonious factor model. That is, the goal is to develop a scale which is not excessively lengthy and in which the items are strongly related to a small number of factors. The factor solution can be used to guide the elimination of items which do not load strongly on any factor or items which ambiguously load on two or more (DeVellis, 2003). Items with factor loadings below .70, .50. or .30 can be deleted. For exploratory factor analysis, the lowest threshold may be desirable to avoid deletion of items which may ultimately be of value. Additionally, items with cross-loadings which are less than .15 different from the from an item's highest factor loading may be considered for deletion. Generally, it is preferable to eliminate items based on cross-loadings only after the factor solution has been determined; it is possible that an item with a relatively high cross-

loading could be retained if the other factor upon which it loaded is deleted or collapsed into another factor.

Additionally, factors with fewer than three items are frequently deleted as this is not considered a stable factor structure. Finally, after rotation, item communalities can be used to assess the proportion of variance accounted for by the factors. Therefore, items with low communalities (less than .40) which are not highly correlated with one or more factors may be appropriate for deletion (Worthington & Whittaker, 2006).

Items which do poorly both in terms of reliability and validity are obvious candidates for removal. Beyond that, it is up to the researcher to look carefully at all the results and to make decisions which will contribute to the scale's overall psychometric properties and usefulness in further research.

Current Study: Method

The purpose of this study was to develop a valid scale for use in operationalizing the behaviors of child welfare supervisors for future studies. Each item is intended to be related to the actual behaviors of supervisors as perceived by their supervisees. The method outlined below utilized the practices and recommendations described in the literature above to guide the research project to develop the Experience of Supervision Scale.

Research questions and hypotheses. Specifically, the study addressed the following research questions and hypotheses:

1) To what degree is the Experience of Supervision Scale a reliable measure of child welfare workers' perceptions of supervisors' behaviors?

    H1: The Experience of Supervision Scale will possess reliability as evidenced by an adequate Cronbach's alpha among items that purport to

measure each of the casework supervision model's three supervisory functions.

2) To what degree is the Experience of Supervision Scale a valid measure of child welfare workers' perceptions of supervisors' behaviors?

H1: The Experience of Supervision Scale will possess validity as evidenced by face validity.

H2: The Experience of Supervision Scale will possess validity as evidenced by content validity.

H3: The Experience of Supervision Scale will possess validity as evidenced by a simple factor structure comprised of three factors which correspond to the casework supervision model's three supervisory functions.

H4: The Experience of Supervision Scale will possess construct validity as evidenced by a divergent relationship between the Experience of Supervision Scale and demographic data provided by participants

Operational definitions. The Experience of Supervision scale will be based upon the theoretical framework of Kadushin's casework model of supervision. The three functions of supervision as defined by Kadushin will provide the basis for item development and validity will be assessed based on the degree to which the draft scale conforms to the theoretical model.

Administrative supervision is operationally defined as those supervisory practices concerned with "the correct, effective, and appropriate implementation of agency policies and procedures" (Kadushin & Harkness, 2002, p. 20).

Educational supervision is defined as those supervisory practices intended to improve worker skills and capacities; the primary goal of educational supervision is "to dispel ignorance and upgrade skill" (Kadushin & Harkness, 2002, p. 20).

The primary goal of supportive supervision is "to improve morale and job satisfaction" (Kadushin & Harkness, 2002, p. 20). Supportive supervision is defined as those supervisory practices which help the supervisee feel good about his or her job.

Kadushin's model is based on the theoretical assumption that the supervisory roles complement each other and that "the good supervisor" is one who is skilled and comfortable using his or her skills in all three domains. Additionally, the three supervisory functions complement each other so that a "good" supervisor is one that performs all three functions well (Kadushin & Harkness, 2002).

Item generation. The goal of this study was to develop a scale that 1) is consistent with nominal definitions of the functions of supervision, 2) has distinguishable dimensions, and 3) accounts for the major components of child welfare supervision. The deductive method of item generation was used. The casework model of supervision developed by Kadushin provided the conceptual dimensions.

The first stage of this research project utilized qualitative data gathered from child welfare workers and child welfare supervisors about supervisory behaviors to generate a list of potential item stems. During the spring of 2009, child welfare workers and child welfare supervisors from a small group of local agencies in the Central and Piedmont regions of the Commonwealth of Virginia were invited to participate in focus groups in order to facilitate the development of a scale to measure supervisory behaviors. The localities were selected primarily because they had a sufficiently large child welfare staff to support a worker focus group and to

have multiple child welfare supervisors. Additionally, at each agency there was a contact person with whom the research had previously developed a professional relationship and who was willing to facilitate the planning for a focus group.

Two focus groups for child welfare workers only were held in two agencies; between both focus groups a total of 12 child welfare workers participated. While focus groups with child welfare supervisors were planned at two additional local agencies, difficulties around individual supervisors' schedules resulted in a series of small interview groups, rather than focus groups, being conducted. In one agency a group of three supervisors met with the researcher, and then another two additional supervisors were interviewed together at a later date. In a separate agency, a small group interview was scheduled with three supervisors, although ultimately, only two participated. In total, 7 child welfare supervisors and 12 child welfare workers participated in the item generation process.

The focus group process was designed with attention to issues around confidentiality, privacy, informed consent, and voluntary participation. A standard consent form required by the Virginia Commonwealth University IRB (Appendix A) was used to obtain consent. Cover sheet information included: the purpose of the study, the voluntary nature of participation; the importance of the participant's input and a review of the risk related to sharing, as confidentiality could not be fully guaranteed. Focus group members were specifically instructed to respond to questions with information about supervision "in general" rather than about any individual's supervisor in particular. Additionally, the researcher prefaced discussion with a reminder that information shared in the focus group should remain confidential and that no one was required to share anything that made him or her uncomfortable and that no one was required to participate. The focus group participants were specifically warned about the risk that another group member

may interpret something shared as particular to the participant's supervisor and share that information outside the focus group. All participants were directly encouraged to maintain confidentiality.

The focus group schedule included 3 categories of questions (Appendix B). First, participants were asked to develop a list of all the things that supervisors do. Next, participants were provided with the definition of a category of supervisory behaviors as delineated by Kadushin's casework supervision model. As each definition was presented, participants were asked to add supervisory activities that came to mind related to that category. After administrative, supportive, and educational supervisory behaviors were addressed explicitly, the group was again asked a general question about supervisory behaviors and encouraged to add anything that might have previously been left out.

In each group all questions were used, although the amount of time spent answering each question varied greatly. Focus groups lasted between 60 and 90 minutes. Refreshments were provided for participants of the focus groups with child welfare workers. Refreshments were offered to the child welfare supervisors, but when the interviews were scheduled, each contact person declined to have refreshments provided.

Focus groups were not taped or transcribed, but the researcher did take notes during the group. No identifying information was collected from the participants, beyond their signed consent forms, but these forms were never associated with any data collected. The researcher's notes did not include identifying information of any kind. The notes taken were used to develop potential item stems for the Experience of Supervision Scale.

Selection of scale format. While the focus groups were underway and distinct ideas for use as items were being extracted from the focus group data, efforts were also underway to

determine the most appropriate scale format for use for the Experience of Supervision Scale. The goal of this study was to produce a scale which could be used in complex research regarding child welfare workers' performance and retention. In order to meet this goal, the Experience of Supervision Scale results would necessarily produce at least interval data in order to be appropriate for use in multivariate statistical analysis (DeVellis, 2003; McDonald, 1999; Rubin & Babbie, 2001).

This researcher ultimately settled on a Likert scale format. While Likert scales actually produce ordinal data, they are routinely used with statistical tests for interval measures, with good results (Bardo, Yeager & Klingsporn, 1982). Additionally, Likert scales are considered ideal for measuring opinions, beliefs, and attitudes (DeVellis, 2003; Rubin & Babbie, 2001), which is consistent with the intended use of the Experience of Supervision Scale. Finally the format is one that is regularly used in research concerning child welfare workforce issues, increasing the compatibility of the new scale with measures already being used.

Selection of response set format. As previously noted, Likert scales can include response options ranging from 3 to 11 choices. However, there is not one option that is clearly better than the others. Internal consistency has been found to increase with greater response options up to 5 choices; additional choices yielded no additional benefit. More response options have been found to contribute to greater central tendency error and a reduction in variance (Bardo & Yeager, 1982; Bardo, Yeager & Klingsporin, 1982). Five and seven point Likert response sets have been found to function similarly when compared (Dawes, 2008; Rhodes et al., 2010) Hinkin (1995) recommends use of a 5 or 7 point Likert format as adequate for yielding good internal consistency without being unduly complicated or lengthy for the respondent.

Ultimately, this researcher settled upon a 5 point response format as there was no compelling evidence that another choice would be better. Simpler was chosen over more complex, with consideration for how the respondents might perceive the relative ease with which the questionnaire might be completed. As the item stems were behaviorally based, response choices based on frequency (e.g. never/always) rather than opinion (e.g. agree/disagree) were selected as the better conceptual fit (Spector, 1992). The end point and the mid-point of the response set were labeled: never, 50% of the time, and all the time. The two additional points were left blank, as no labels were considered to result in more equal intervals than leaving those response choices blank.

Respondents were not provided with an opportunity to select "Don't know" in order to maximize meaningful responses to the selected items. However, respondents were given the option of skipping items; if particular items were frequently skipped, they could be eliminated as "bad" items prior to further data analysis.

Refining the item pool. The second stage of the project was the use of an expert panel to review item stems for readability and clarity and to do an initial assessment of both face and content validity relative to the three proposed subscales. Twelve experts were asked to review the items and provide feedback to the researcher about them. The expert panel was comprised of child welfare workers (n=4); child welfare supervisors (n=5); and experienced researchers with expertise in the area of child welfare and/or scale development (n=3).

The item pool was shared with the panel via an internet based survey set up in Surveymonkey.com. Panel members were asked to note any items which were unclear or confusing and also to sort each item into a category of supervisory function based on Kadushin's case work supervision model. The results of this process are reported in Chapter Four.

Questionnaire development.

***Questionnaire design*.** Once the items, item format, and response format for the Experience of Supervision Scale were selected, consideration was given to the questionnaire to be used in administration of the instrument. As previously discussed, a web-based survey method was to be used. Surveymonkey.com is a relatively inexpensive, readily available and easy to use web-based survey software. Surveymonkey.com was initially used by this researcher to facilitate the Expert Panel's review of the initial item pool, with good results. Additionally, use with the Expert Panel did not produce any issues around internet security or software compatibility, which can be a significant problem with web-based surveys (Couper, 2000).

Additional consideration was next given to the design of the questionnaire within the constraints of the Surveymonkey.com software. Galesic, Tourangeau, Couper and Conrad (2008) measured the amount of time that survey participants actually looked at response options utilizing sophisticated eye-tracking equipment with study participants responding to questions in a web-based survey. They found that respondents spend less time reading options which are not readily available (hidden in a drop down list) and in the second half of a list of options. Their recommendations include "making the questionnaire short and interesting" and making all the response options immediately available.

The Likert 5 point response format fit within the normal Surveymonkey.com page, permitting a Portrait orientation. Items were entered with page breaks so that the respondents proceeded to the next page to answer additional questions, with only minimal scrolling down the page. Radio buttons were utilized instead of providing text boxes in which participants could enter a selection or boxes containing drop-down choices. The goal was to improve ease and

speed of completing the questionnaire by using a mouse-only questionnaire format; respondents were not required to type to complete the survey (Couper, Traugott, and Lamias, 2001).

Positively and negatively worded items. The use of inversely worded items is a strategy which is recommended to reduce or at least to assess the degree to which acquiescent response set bias may influence the scores of a scale (Spector, 1992). However, interpreting the results of scores on items which have been inverted can be challenging. Particularly in an untested scale, how can it be determined if a high or low score on an inverse item is the result of response set bias or a product of a poorly designed item (DeVellis, 2003)? This researcher made the decision to use only positively worded items based upon the assessment that for a first attempt to validate a scale it was preferable to have as straightforward an evaluation of validity and reliability as possible.

Inclusion of additional items. The data analysis plan for this study included use of factor analysis to evaluate the fit of the scale to the underlying three supervisory functions theoretical model. However, consideration was also given to whether additional items to be used in construct validity should also be included in the questionnaire. Construct validity requires that a hypothesis by developed about the relationship between the construct of interest and another construct. As discussed previously, while child welfare supervision has been identified as an important variable in a number of studies, as a construct, it has not been well defined nor consistently measured. When construct validation is conducted concurrently with theory validation and variables do not behave as hypothesized, it is unclear whether the results reflect construct invalidity or scale invalidity (Spector, 1992).

Additionally, there are concerns about contamination of responses due to exposure to another scale and excessive length. For these reasons, it is not recommended that an additional

scale or scales be included in an initial scale development administration (DeVellis, 2003; Worthington & Whitaker, 2006).

For the purposes of this study, this researcher decided that rather than add to the length of the questionnaire with a second scale or additional items to measure an additional construct for which there was no strong theoretical relationship, an initial assessment of construct validity would be estimated through evaluation of the relationship of the scale scores with demographic variables.

Demographic variables. As the purpose of the study is to test the psychometric properties of the proposed Experience of Supervision Scale, rather than generalizable study results, demographic questions were included primarily as controls, to insure that those responding to the survey were intended participants. Secondly, items were chosen to permit a description of the study sample and to facilitate comparison between the sample and the population. The Virginia State Department of Social Services conducted a brief state-wide survey regarding child welfare workers' attitudes towards family engagement during the months that the questionnaire for this study was being developed. The state does not have readily accessible demographic information for the child welfare worker population; demographic questions for this questionnaire were modeled after the demographic questions included in the family engagement survey. At that point, inquiries about obtaining the results of the family engagement survey's demographic questions were positively received. This researcher's intention was to compare the description of the Experience of Supervision scale sample to the (presumably) larger family engagement survey sample. However, to date, only summary data (which is not comparable) has been made available to this researcher.

Demographic variables collected for this study included job category, program area, length of employment in the field, gender, age range, level of education, and whether the respondent's degree was in social work.

Sample frame formulation. The final stage of the project was to administer the draft Experience of Supervision Scale to a sample of child welfare workers. The research plan was submitted to the Virginia Commonwealth University IRB review board; this portion of the study was determined to be an exempt study.

A purposive, non-random sample of child welfare workers from cities and counties across Virginia was developed. The questionnaire was developed into a web based survey through Surveymonkey.com. (Appendix C). Because Virginia's Department of Social Services is a locally administered program, not all of the child welfare workers in the state are state employees. No master list of all child welfare workers and their contact information exist in any data base. It is not possible therefore to develop a sampling frame from which to randomly select potential study participants. Instead, purposive sampling was utilized to try to obtain as many responses as possible, while attempting to insure that child welfare workers from large and small local agencies, rural, and urban communities and from all five geographic regions in the state were included in the study.

In the spring of 2010, every local agency director in Virginia received an e-mail detailing the purpose and methodology of the project and requesting that a contact person be established at that agency. Several weeks later, the e-mail was re-sent to each director who had yet to respond (Appendix D).

Once a contact person was established, he or she was sent an email again detailing the purpose and methodology of the project (Appendix E) which was followed up by an email

containing an invitation to the child welfare workers of that agency to participate in the study and a direct link to the web based survey in Surveymonkey.com (Appendix F). This email was designed to be forwarded "as is." After a number of weeks, the contact person was contacted again and asked to forward a reminder email to the child welfare workers. The reminder email was also sent to the contact person, again designed to be forwarded "as is" and again containing the web link to the survey (Appendix G).

This strategy allowed for complete anonymity to be preserved for the study participants. The researcher never received potential or actual participants' names or contact information from the agency.

In all of the emails sent regarding the study, recipients were informed that six $50 Best Buy gift cards were to be selected from all study participants as incentives for participating. At the end of the web based survey, participants were directed to a separate "survey" also set up through Surveymonkey.com where they were able to enter an email address to permit this researcher to both potentially select her or him as a gift card recipient and also to contact her or him to make arrangements to provide the gift card. The contact information provided was retained entirely separately in Surveymonkey.com so that there was no way to associate the provided email addresses with any individual responses. Nor was it necessary to utilize a state email address which could identify the worker's name and/or locality (e.g. parente@charlottesville.org); many participants chose to provide gmail, hotmail or other personal email account addresses.

Three participants were randomly selected through use of a random number generator mid-way through the data collection period to receive the first 3 gift cards. Several weeks after final reminders were sent to all the designated contact persons, a final three gift card recipients

were selected. All but one of the gift cards were mailed; when contacted via email regarding having been selected the study, five selected participants provided their addresses. The final gift card was picked up from this researcher by the selected participant, who identified himself as living in the vicinity of the researcher's office.

Chapter 4: Results

This chapter presents the results of a multi-stage research project with the goal of developing and evaluating the psychometric properties of a scale to measure supervisory behaviors from the perspective of the child welfare worker. A review of the stages of the project will be presented including sample characteristics. A detailed report of the findings of factor analysis procedures, an interpretation of factor structure, a review of an assessment for reliability and a first effort to assess concurrent validity will also be presented. Finally, the initial reliability and validity estimations will be incorporated into a rationale for approaching item elimination and refinement of the proposed Experience of Supervision scale.

Initial Item Development

Ultimately the material collected from participants of the focus groups discussed in detail in Chapter Three yielded 172 distinct ideas about supervision (Appendix H). The goal of this project was to develop a scale to measure supervisory behaviors. Therefore, this researcher reviewed the initial list of items and eliminated items which were not consistent with this goal. The initial list of potential item stems was reduced to 108 items stems first by eliminating items which were not expressed as empirical behaviors (e.g. ….. is trustworthy and ….is professionally educated). Secondly, ideas that were similar (e.g. ….establishes deadlines, …..reminds me of deadlines and ….holds me accountable for completing my work on time) were consolidated into a single item (e.g. …..holds me accountable for completing my work on time.)

The remaining 108 item stems were also sorted by the researcher into the three categories of supervisory functions as defined by Kadushin in order to assess the degree to which all three categories were represented in the refined item pool. Thirty-seven administrative, 38 supportive and 33 educational items were identified for inclusion in the item pool to be reviewed by the expert panel (Appendix I).

Face and Content Validity Assessments

The researcher's efforts to refine the item pool described above were the first steps in attempting to establish face and content validity. Face validity is the assessment of the degree to which items appear to measure the construct intended. Items were initially reviewed to determine whether they all appeared to be measuring the construct of supervisory functions or behaviors. For example, items which described character traits associated with good supervisors were deleted, while behaviorally operationalized items were retained.

Then, the researcher sorted the items into the three categories of supervisory behavior using the conceptual definitions from the casework model of supervision for administrative, supportive and educational functions. Content validity is the assessment of the degree to which the items describe the construct in its entirety. Although it was not possible to determine whether all of the items necessary to cover the complete domain of supervisory behaviors were included, the pool of item stems was fairly evenly divided among the three categories of supervisory functions. The items appeared to have content validity relative to the structure of the conceptual model.

The 108 retained items were next reviewed by the expert panel. The panel was asked to sort the items into the three supervisory behavior categories. The survey included the following definitions at the top of each page:

Administrative supervision is operationally defined as those supervisory practices concerned with "the correct, effective, and appropriate implementation of agency policies and procedures."

The primary goal of supportive supervision is "to improve morale and job satisfaction." Supportive supervision is defined as those supervisory practices which help the supervisee feel good about his or her job.

Educational supervision is defined as those supervisory practices intended to improve worker skills and capacities; the primary goal of educational supervision is "to dispel ignorance and upgrade skill."

Panel members were also given the opportunity to select "can't decide," for items which did not appear to them to sort into one of the theoretical categories. These "judges" evaluated each item's fit with the nominal definitions. Items for which multiple experts selected "can't decide" were to be eliminated as having low face validity (not being a good "fit" with the theoretical model defining the construct to be measured).

The process of classifying items described by Hinkin (1995) was also used to assess content validity. Experts were asked to classify the items into the three theoretical dimensions of the scale: administrative, supportive and educational supervisory functions. Items which were not consistently categorized at least 75% of the time, were to be eliminated. The process was designed to increase the likelihood that there would be "good" items across the domain of casework supervisory behaviors.

Following review by the expert panel, 47 items were eliminated because less than 75% of panel members (8 of 12) agreed on the categorization of the item (Hinkin, 1995). Four of these items were sorted as "can't decide" by multiple panel members; they would have been eliminated due to concerns about face validity, if they had not been eliminated through failing to sort consistently into one of the categories of supervisory behaviors. None of the items were eliminated from the item pool solely due to concerns about language; all items which were noted to be confusing also failed to sort reliably and were eliminated on those grounds.

Following the expert panel review, 61 potential item stems remained: 14 administrative supervision items; 28 supportive supervision items; and 19 educational supervision items. Seven items were consistently sorted by the expert panel into a different category of supervisory function than the category original designated by the researcher (Appendix J). These items were retained and re-designated into the category indicated by the expert panel for use in the draft Experience of Supervision scale.

Sample Characteristics

The study design was intended to maximize the number of respondents by carefully protecting their anonymity. Because designated contact persons forwarded the researcher's email invitation to potential respondents, there is no record of how many potential participants received the invitation. Therefore, a response rate for the actual survey cannot be calculated. However, it is known that 1855 child welfare workers were active in the statewide data reporting system as of January, 2011. The final sample (described further below) included 165 participants or 9% of the total population.

Of 122 local agencies contacted, 34 (28%) eventually identified a contact person who received the information to send to the child welfare workers at that agency. There were

responding agencies from all 5 geographic regions of the state and large and small as well as urban and rural agencies were well represented (see Table 1). Because the larger agencies employee so many more child welfare workers per agency, it is very likely that there are more workers in the sample from large urban agencies than there are from small or very small urban areas. However, this is believed to also be the case for the population of child welfare workers in the state.

Although efforts were made to keep the survey very short, several directors initially reported interest, and then later declined to forward the information to child welfare workers citing concerns about workers' current work load (n=4). Several other directors declined to proceed due to recent turnover of a supervisor, staff or manager (n=2). Most frequently, however, the director simply never responded to the researcher's emails.

During the data collection period, 242 individuals began the web-based survey. Of those, 228 completed the survey. However, while the invitation to participate in the survey clearly indicated that child welfare workers were the intended study participants, when asked about their job category at the end of the survey, 25 respondents indicated that they were child welfare supervisors. An additional 2 participants indicated that they were administrative support personnel. These 27 cases were eliminated, leaving 201 surveys completed by child welfare workers for the study. Of those 201 cases, an additional 36 cases were eliminated because they contained some missing data. Factor analysis requires complete cases, and the most conservative approach to handling missing data was utilized for this project. The final sample included 165 cases in which participants responded to all 61 items which comprised the Experience of Supervision scale within the questionnaire.

Table 1

Participating Localities and their Characteristics

Name	Region	Size	Type
Amelia	Central	very small	rural
Chesterfield	Central	mid-size	rural
Fluvanna	Central	small	rural
Goochland	Central	small	rural
Hanover	Central	small	rural
Henrico	Central	mid-size	urban
King and Queen	Central	very small	rural
Richmond	Central	large	urban
Accomack	Eastern	very small	rural
Chesapeake	Eastern	mid-size	urban
Dinwiddie	Eastern	small	rural
Hampton	Eastern	small	urban
James City	Eastern	small	rural
Newport News	Eastern	large	urban
Portsmouth	Eastern	large	urban
Southampton	Eastern	very small	rural
Suffolk City	Eastern	small	rural
Alexandria	Northern	mid-size	urban
Arlington	Northern	mid-size	urban
Culpeper	Northern	small	rural
Fairfax	Northern	large	urban
Fredericksburg	Northern	small	urban
Greene	Northern	very small	rural
King George	Northern	small	rural
Loudon	Northern	mid-size	urban
Stafford	Northern	small	urban
Albemarle	Piedmont	mid-size	rural
Charlottesville	Piedmont	mid-size	urban
Danville	Piedmont	small	urban
Henry-Martinsville	Piedmont	mid-size	rural
Roanoke City	Piedmont	large	urban
Floyd	Western	very small	rural
Grayson	Western	small	rural
Tazewell	Western	mid-size	urban

Most respondents were female (93%) and the most frequent age range indicated was 20 to 30 years old (30%). Fifty-eight percent had a Bachelor's degree. Thirty-nine percent had a Master's degree. Of these, 42% percent indicated that their degree was in Social Work. Respondents identified their program areas predominately as Child Protection Services (37%) and Foster Care (38%). The majority of respondents (36%) had worked in the child welfare field for one to five years (see Table 2).

Although the sample was purposive rather than random, the sample characteristics are similar to the known characteristics of the child welfare worker population for the state. The Virginia statewide child welfare data base (OASIS) requires that education level be entered when new workers are given access to the system. In January 2011, there were 1855 child welfare workers active in the OASIS system. Of these 25% had Masters degrees and 46% had a degree in social work (M. Wade, personal communication, January 2011).

Data Analysis

Preliminary analyses.

Missing data. The frequency distributions for responses for all 61 items were first reviewed for missing data. As respondents were able to skip items, data was reviewed to see if any items had a significant number of non-responses as this would be an indicator that the item was confusing or challenging for respondents. However, while many items had some non-responses, no item exceeded 5 non-responses within the original 201 child welfare cases retained (see Table 3). Therefore, no items were eliminated as bad items based on non-response.

Table 2

Characteristics of Sample Compared to Population

Characteristic	n	%	Statewide data n	Statewide data %
Total (n)	165	100	1855	100
Program area				
Prevention/Family stabilization	20	12.1	—	—
Child Protection Services	61	37.0	—	—
Foster Care	63	38.2	—	—
Adoption	6	3.6	—	—
Other (specialized titles)	11	6.7	—	—
Years of experience				
Less than 1	12	7.3	—	—
1 to 5 years	59	35.8	—	—
5 to 10 years	42	25.5	—	—
More than 10	49	29.7	—	—
Education level				
High school	3	1.8	54	3
2 year degree	0	0.0	11	1
Bachelor's degree	95	57.6	1326	71
Master's degree	64	38.8	461	25
Social Work degree				
Yes	69	41.8	849	46
No	93	56.4	1006	54
Gender				
Male	15	9.1	—	—
Female	145	87.9	—	—
Age range				
20 to 30	60	36.4	—	—
31 to 40	41	24.8	—	—
41 to 50	28	17.0	—	—
51 to 60	30	18.2	—	—
61 or older	3	1.8	—	—

Table 3

Descriptive Statistics for Items

Item	Missing data (n = 201)	*M* (n = 165)	*SD*	Minimum	Maximum
Q1	0	3.82	1.137	1	5
Q2	0	4.31	.888	2	5
Q3	0	4.25	.933	2	5
Q4	1	3.85	1.172	1	5
Q5	0	3.18	1.352	1	5
Q6	2	3.66	1.266	1	5
Q7	0	3.59	1.348	1	5
Q8	1	3.72	1.287	1	5
Q9	0	4.01	1.096	1	5
Q10	0	3.75	1.223	1	5
Q11	0	3.84	1.215	1	5
Q12	1	4.01	1.205	1	5
Q13	0	3.67	1.231	1	5
Q14	1	4.14	1.189	1	5
Q15	0	3.95	1.282	1	5
Q16	2	2.08	1.332	1	5
Q17	2	4.12	1.098	1	5
Q18	0	3.98	1.131	1	5
Q19	0	3.99	1.184	1	5
Q20	0	4.22	1.078	1	5
Q21	2	4.01	1.202	1	5
Q22	1	3.82	1.216	1	5
Q23	2	4.05	1.103	1	5
Q24	3	3.68	1.334	1	5
Q25	0	4.12	1.128	1	5
Q26	0	4.12	1.141	1	5
Q27	0	4.08	1.123	1	5
Q28	0	4.43	.885	1	5
Q29	0	4.12	1.217	1	5
Q30	0	4.04	1.226	1	5
Q31	1	4.17	1.010	1	5
Q32	1	4.10	1.102	1	5
Q33	2	3.19	1.309	1	5

Note. Continued next page

Table 3 (continued)

Descriptive Statistics for Items

Item	Missing data (n = 201)	*M* (n = 165)	*SD*	Minimum	Maximum
Q34	1	4.28	1.045	1	5
Q35	1	3.82	1.199	1	5
Q36	1	3.22	1.454	1	5
Q37	1	3.70	1.250	1	5
Q38	5	3.01	1.660	1	5
Q39	3	3.46	1.285	1	5
Q40	2	4.52	.874	1	5
Q41	2	3.96	1.131	1	5
Q42	2	4.18	.919	2	5
Q43	2	3.49	1.286	1	5
Q44	2	3.95	1.144	1	5
Q45	2	4.64	.797	1	5
Q46	2	2.93	1.500	1	5
Q47	4	3.48	1.383	1	5
Q48	2	4.07	1.102	1	5
Q49	3	4.19	.943	1	5
Q50	2	4.05	1.029	1	5
Q51	4	3.73	1.265	1	5
Q52	3	3.64	1.330	1	5
Q53	1	3.99	1.118	1	5
Q54	4	3.92	1.197	1	5
Q55	1	4.27	.995	1	5
Q56	6	3.76	1.132	1	5
Q57	2	3.30	1.332	1	5
Q58	5	3.85	1.243	1	5
Q59	3	3.71	1.307	1	5
Q60	2	3.50	1.277	1	5
Q61	1	3.91	1.244	1	5

Univariate analysis. All cases with missing data within the 61 items of the Experience for Supervision scale were then deleted (consistent with constraints for Factor Analysis). The frequency distribution for each item was then reviewed for the remaining 165 cases. Table 3

displays the summary statistics for each item. Generally responses tended to cluster towards the more positive/ higher scoring response options. All but two items had a mean above the midpoint (3). However, examination of the frequency distributions shows that while there is a slight positive skew for many items, the means are a reflection of scores spread across the response options, averaged to the middle score. For all but 3 items (Q2, Q3, and Q42), respondents selected the minimum and maximum categories. For these three items, 4 of the 5 categories were selected.

The standard deviation provides an index of variability when variables are measured on an interval or ratio scale. It is the most stable measure of variability because the standard deviation takes into account every score in the distribution. A large standard deviation indicates that the distribution is spread out away from the mean. A small standard deviation indicates that the distribution is clustered more closely around the mean. For these data, standard deviations tended to be modest, ranging from a minimum of .797 for Q45 to a maximum of 1.660 for Q38. The frequency distributions and the standard deviations for the items suggest that these data have good variability.

Variability of items is generally a desired feature of a scale. Variance is useful in analytic procedures intended to detect subgroups of respondents that may be of interest. Very high or very low means may be an indicator of lack of variability, or may be a function of the sample. Very skewed distributions of an item would also be of concern as this could be an indicator of lack of variation on an item. However, for these data, frequency distributions did not suggest that any item be eliminated based on the distribution of responses on the univariate level.

Assessment of multivariate assumptions for factor analysis. Data were next screened for the presence of outliers. Outliers are of concern with both factor analysis and regression as

unusual values can result in distortions of the findings. Cook's D was used to produce a measure of the impact of an observation on the estimated model. There are several thresholds which are recommended for use in identifying large Cook's D values (see Table 4).

Table 4

Calculated Value of Criterion

Criterion	Calculated value
$D > 4/(n - k - 1)$	$4/(165 - 6 - 1) = .025$
$D > 1$	1
$D > 4/n$	$4/165 = .024$
Average of above criteria	.3497

When the average for the three suggested criterion was utilized with the total scale score and the three subscale totals as DVs, no cases were identified as outliers. Outliers were not found to be problematic in these data.

Homoscedasticity is the assumption that the variability in scores for one variable is roughly the same for all values of all other variables in the model. The absence of homoscedasticity, which indicates heteroscedasticity, can result in parameter estimates that are too large or too small. Results can be overly optimistic (type I error) or overly pessimistic (type II error) as a result. One approach to assess data for heteroscedasticity is to examine the plots of standardized predicted values as a function of standardized residuals values. In these data, the normal P-P plot and the scatterplot of the standardized predicted values as a function of standardized residual values, demonstrate a low amount of heteroscedasticity; the data have a

slightly positive skew. While homoscedasticity is assumed, Factor Analysis is robust for moderate violations of homoscedasticity.

Factor analysis procedure

Factor extraction, factor rotation, and initial solution. A series of exploratory factor analyses using SPSS (PASW) 17 were conducted in this study. These analyses sought to systematically combine the following three dimensions: 1) principal components factor analysis versus principal axis extraction 2) varimax (orthogonal) versus direct oblimin (oblique) rotations and 3) number of factors (two versus three based on Kadushin's theoretical model) (see Table 5).

Table 5

Exploratory Factor Analyses Conducted on Experience of Supervision Scale Data

Principal components extraction	Principal axis extraction
varimax rotation, three factors	varimax rotation, three factors
varimax rotation, two factors	varimax rotation, two factors
direct oblimin rotation, three factors	direct oblimin rotation, three factors
direct oblimin rotation, two factors	direct oblimin rotation, two factors

The initial principle components factor analysis (PCFA) with varimax rotation, an orthogonal rotation, produced a three factor solution consistent with the theoretical model being tested. However, a PCFA analysis with an oblique rotation, direct oblimin, was also run, as the theoretical model suggests that the factors in the model may be correlated. Additionally, a two factor solution was developed using both the orthogonal and oblique rotation extraction methods. Finally, principal axis factoring (PAF) was utilized with a three factor solution, two factor solution and both orthogonal and oblique rotation methods. While the oblique rotation solution

with PCFA found that the factors were modestly correlated, the oblique rotation solution did not produce significant differences from the original PCFA with varimax rotation three factor solution. The PAF models were also very similar.

The two factor model in all four instances (PCFA with orthogonal rotation extraction, PCFA with oblique rotation extraction, PAF with orthogonal rotation extraction, and PAF with oblique rotation extraction) resulted in a model that was harder to interpret and weaker in terms of the percentage of variance explained. Because the factor solutions were very similar and because the principal components factor analysis with varimax rotation are commonly used factor extraction and factor rotation strategies with well known properties, the principal components factor solution with varimax rotation and a three factor structure was used for scale refinement and validity assessment.

The Kaiser-Meyer-Olkin (KMO) test provides a measure of the factor analyzability of a set of variables. This measure varies between zero and one, and values closer to one are better. A value of 0.60 is a suggested minimum. A low KMO would indicate low factor analyzability. These data produced a KMO of .943, which is considered strong.

The principal components factor solution with varimax rotation resulted in three factors which explained 57.7% of the variance in the model. The variance accounted for was fairly evenly distributed across the three factors extracted. Additionally, each factor had multiple items (more than 2) with factor loadings greater than .50 (see Table 6).

Table 6

Total Variance Explained for Exploratory Factor Analysis with Varimax Rotation of Experience of Supervision Scale

	Initial Eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
Component	Total	% Variance	Cumulative %	Total	% Variance	Cumulative %	Total	% Variance	Cumulative %
1	30.819	50.523	50.523	30.819	50.523	50.523	13.403	21.973	21.973
2	2.398	3.931	54.454	2.398	3.931	54.454	11.843	19.415	41.387
3	2.001	3.280	57.734	2.001	3.280	57.734	9.971	16.346	57.734

Note. Extraction Method: Principal Component Analysis.

This solution's communalities, factor scores and factor loadings are reported in Appendix K. Communalities indicate the degree to which an item contributes to a factor. Values range from zero to one. The closer the score is to one, the greater the variance of the variable accounted for by the factor structure (Mertler & Vannatta, 2005). The proportion of variance explained by the factor model is very good for some variables (>.70) (e.g. Q19= .722, Q23=.700, Q29=.700, Q30=.709, Q52=.723, Q54=.719) while low for other variables (e.g. Q12= .484, Q16=.403, Q31=.460, Q32=.483, Q38= .353, Q45=.419, Q49=.495, Q50=.432, Q61=.471). No item had a communality of less than .30 and two of the items with the lowest estimates of communality (Q10 and Q35) also loaded on multiple factors and were eliminated when the scale was refined further. The factor model explained more than 50% of the variance in 50 of the 61 items.

Rotated factor loadings are the correlations between the variable and the factor. The higher the factor score of the item the greater the correlation. In an ideal factor structure (simple structure) each item would load perfectly on one factor alone. Factor scores can be used to retain or delete items in order to approximate a simple structure. These data produced an easily

interpreted three factor solution, which was further simplified through the elimination of items which did not contribute strongly to the factor structure.

The factor score threshold of .50 was used to eliminate those items which did not load strongly onto any of the factors. Nine items, Q9, Q10, Q13, Q17, Q22, Q25, Q35, Q44, and Q51, did not load onto any of three factors with a factor score > .50. Additionally, several items loaded very similarly on more than one factor. As recommended by Worthington and Whittaker (2006), these items, which did not produce factor loadings with > .15 difference between the factors, were also eliminated: Q4, Q6, Q8, Q11, Q15, Q18, Q19, Q21, Q23, Q26, Q31, Q32, Q37, Q40, Q43, Q49, Q53, Q54, Q58, and Q59 (see Appendix K). Thirty-two items were retained for further analysis. These included 15 items associated with the first factor, 10 items associated with the second factor and seven items associated with the third factor.

Item Analysis and Reliability Assessment

Next, the retained 32 items were assessed for internal consistency. Cronbach's alpha, a measure of internal consistency, provides a conservation estimate of a scale's reliability; it is the unique estimate of the expected correlation of one test form with a hypothetical alternative form of the same length (Carmines & Zeller, 1979). Cronbach's alpha and alpha by item were used. A good scale alpha will be in the .70 to .80 range (Bland & Altman, 1997; DeVellis, 2003).

A reliability estimate was developed for each subscale (factor) and for the entire Experience of Supervision Scale. The alpha coefficient for the total scale was .962, well above the .70 threshold for internal consistency discussed by Nunnally (1978). The first two subscales also produced an alpha coefficient in the .900 range. The third subscale's alpha coefficient was .878 (see Table 7).

Table 7

Assessment of Reliability for the Refined Experience of Supervision Scale

Total experience of supervision scale (alpha = .962)	Subscale item to total statistics		Total scale item to total statistics	
	Corrected item to total correlation	Alpha if item deleted	Corrected item to total correlation	Alpha if item deleted
Factor 1 (alpha = .938)				
Q33 When assigning cases, is sensitive to the kinds of cases I prefer	.678	.933	.633	.961
Q46 Asks me about what motivates me	.731	.931	.703	.961
Q52 Helps me to recognize when a particular case is really stressing me out	.777	.930	.799	.960
Q39 Advocates for resource development to address resource gaps	.748	.931	.739	.960
Q57 Provides opportunities to observe other areas of practice	.701	.932	.673	.961
Q60 Implements strategies or develops resources to help manage unreasonable caseloads	.736	.931	.736	.960
Q47 Encourages me to take vacation	.625	.934	.607	.961
Q24 Advocates for systems intervention to maintain reasonable caseloads for staff	.738	.931	.735	.960
Q36 Uses observations of my work in the field to help me improve my practice skills	.729	.931	.731	.960
Q16 Uses role play to help me practice new skills	.566	.936	.522	.962
Q56 Provides opportunities for me to try new things	.700	.932	.707	.961
Q7 Encourages me to connect training I have received to specific case situations and/or my intervention efforts	.677	.933	.687	.961
Q61 Suggests trainings I might attend	.647	.933	.659	.961
Q38 Comes to court to support me when I have a challenging court case	.557	.937	.529	.962
Q5 Asks if supervision is meeting my needs	.674	.933	.695	.961
Factor 2 (alpha = .923)				
Q34 Takes the time to understand my side of the situation when there is a complaint	.733	.914	.639	.961
Q2 Is available for consultation when I have a case crisis	.735	.914	.640	.961

Note. Continued next page

Table 7 (continued)

Assessment of Reliability for the Refined Experience of Supervision Scale

Total experience of supervision scale (alpha = .962)	Subscale item to total statistics		Total scale item to total statistics	
	Corrected item to total correlation	Alpha if item deleted	Corrected item to total correlation	Alpha if item deleted
Factor 2 (alpha = .923) (continued)				
Q29 Provides a safe place to talk about feeling overwhelmed	.783	.911	.757	.960
Q14 Takes an interest in me as a person	.772	.912	.748	.960
Q30 Facilitates good team work	.800	.910	.785	.960
Q20 Pitches in and helps handle emergencies	.734	.914	.710	.961
Q42 Is available to me when I have a problem	.754	.913	.709	.961
Q3 Models appropriate personal/ professional boundaries	.678	.917	.646	.961
Q50 Is accepting when I am not able to get everything done on time	.569	.922	.519	.962
Q45 Supports taking time off to deal with family emergencies	.535	.923	.493	.962
Factor 3 (alpha = .878)				
Q27 Monitors progress towards deadlines	.755	.844	.616	.961
Q55 Holds me accountable for completing my work on time	.619	.863	.494	.962
Q28 Consults with specialists or attorney when clarification regarding policy is needed	.685	.859	.577	.961
Q41 Filters policy and practice changes so I get exactly the information I need to do my job	.640	.860	.628	.961
Q12 Critiques my documentation to improve quality and completeness	.617	.864	.536	.962
Q48 Strengthens collaborative relationships with community partner agencies	.660	.857	.681	.961
Q1 Works strategically to improve efficiency within the agency	.663	.857	.721	.961

All individual items had high correlations to the subscale with which they were associated. Most items also had high correlations with the entire scale. Two items had

correlation coefficients less than 0.5 with the entire scale. These items were considered for deletion, but ultimately retained as they were not contributing negatively to the reliability estimates. No items were deleted based on the reliability assessment. The high alphas derived here indicate an acceptable level of reliability for the revised Experience of Supervision scale.

Interpreting the Factors

After the scale had been refined utilizing information from the factor analysis and reliability was estimated for the subscales and total scale, the Experience of Supervision scale was a three factor scale comprised of 32 items. The next step was to use the item loadings for the items which had been retained to interpret the underlying factor structure.

A second principle component factor analysis (PCFA) with varimax rotation was conducted to evaluate the degree to which the refined scale might produce factor loadings which were more consistent with Kadushin's model of casework supervision. Communalities suggest that the derived three factor model explained more than 50% of the variance for all but five of the 32 items. The factor analysis for the refined scale produced a three factor solution with strong and unique loadings (See Table 8). All of the items continued to load strongly onto the factor with which they were originally associated. However, several items changed position in terms of the strength of the loading.

The individual items' factor loading were then evaluated in comparison to the category of supervisory behaviors to which they had been assigned by the expert panel (Table 9). The items which sorted strongly onto the second and third factors were consistent with the casework supervision model's supportive and administrative functions of supervision. For the second

Table 8

Factor Loading for Exploratory Factor Analysis with Varimax Rotation for Refined Experience of Supervision Scale

	Item	Components			Communality
		1	2	3	
Factor 1 (Performance support)					
Q33	When assigning cases, is sensitive to the kinds of cases I prefer	**.720**	.281	.048	.601
Q39	Advocates for resource development to address resource gaps	**.661**	.153	.441	.681
Q57	Provides opportunities to observe other areas of practice	**.681**	.337	.115	.590
Q46	Asks me about what motivates me	**.680**	.260	.249	.592
Q60	Implements strategies or develops resources to help manage unreasonable caseloads	**.647**	.286	.340	.615
Q24	Advocates for systems intervention to maintain reasonable caseloads for staff	**.642**	.245	.387	.622
Q47	Encourages me to take vacation	**.623**	.307	.099	.492
Q56	Provides opportunities for me to try new things	**.616**	.411	.192	.585
Q52	Helps me to recognize when a particular case is really stressing me out	**.613**	.421	.354	.679
Q16	Uses role play to help me practice new skills	**.603**	.059	.222	.417
Q11	Uses observations of my work in the field to help me improve my practice skills	**.596**	.289	.390	.591
Q38	Comes to court to support me when I have a challenging court case	**.585**	.126	.197	.397
Q61	Suggests trainings I might attend	**.560**	.245	.350	.496
Q7	Encourages me to connect training I have received to specific case situations and/or my intervention efforts	**.526**	.320	.370	.516
Q5	Asks if supervision is meeting my needs	**.507**	.331	.395	.523
Factor 2 (Emotional support/ Availability)					
Q34	Takes the time to understand my side of the situation when there is a complaint	.216	**.768**	.207	.679
Q2	Is available for consultation when I have a case crisis	.114	**.738**	.375	.698
Q29	Provides a safe place to talk about feeling overwhelmed	.442	**.657**	.250	.690
Q42	Is available to me when I have a problem	.230	**.657**	.436	.675
Q14	Takes an interest in me as a person	.391	**.649**	.307	.669

Note. Continued next page

Table 8 (continued)

Factor Loading for Exploratory Factor Analysis with Varimax Rotation for Refined Experience of Supervision Scale

	Item	Components			Communality
		1	2	3	
Factor 2 (Emotional support/ Availability) (continued)					
Q20	Pitches in and helps handle emergencies	.314	**.632**	.358	.626
Q30	Facilitates good team work	.404	**.631**	.383	.708
Q50	Is accepting when I am not able to get everything done on time	.396	**.619**	-.100	.550
Q3	Models appropriate personal/ professional boundaries	.171	**.596**	.462	.597
Q45	Supports taking time off to deal with family emergencies	.386	**.564**	-.086	.474
Factor 3 (Administrative activities)					
Q27	Monitors progress towards deadlines	.220	.166	**.795**	.708
Q55	Holds me accountable for completing my work on time	.237	-.016	**.728**	.587
Q12	Critiques my documentation to improve quality and completeness	.199	.151	**.691**	.539
Q28	Consults with specialists or attorney when clarification regarding policy is needed	.107	.358	**.660**	.574
Q41	Filters policy and practice changes so I get exactly the information I need to do my job	.322	.258	**.595**	.525
Q1	Works strategically to improve efficiency within the agency	.371	.366	**.588**	.617
Q48	Strengthens collaborative relationships with community partner agencies	.373	.347	**.533**	.544

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 7 iterations. Factor loadings > .50 are in boldface.

Table 9

Retained Items for Experience of Supervision Scale and Expert Panel Categories

	Revised scale items	Expert panel category
Factor 1 (Performance support)		
Q33	When assigning cases, is sensitive to the kinds of cases I prefer	Supportive
Q39	Advocates for resource development to address resource gaps	Administrative
Q57	Provides opportunities to observe other areas of practice	Educational
Q46	Asks me about what motivates me	Supportive
Q60	Implements strategies or develops resources to help manage unreasonable caseloads	Administrative
Q24	Advocates for systems intervention to maintain reasonable caseloads for staff	Administrative
Q47	Encourages me to take vacation	Supportive
Q56	Provides opportunities for me to try new things	Educational
Q52	Helps me to recognize when a particular case is really stressing me out	Supportive
Q16	Uses role play to help me practice new skills	Educational
Q11	Uses observations of my work in the field to help me improve my practice skills	Educational
Q38	Comes to court to support me when I have a challenging court case	Supportive
Q61	Suggests trainings I might attend	Educational
Q7	Encourages me to connect training I have received to specific case situations and/or my intervention efforts	Educational
Q5	Asks if supervision is meeting my needs	Supportive
Factor 2 (Emotional support/ Availability)		
Q34	Takes the time to understand my side of the situation when there is a complaint	Supportive
Q2	Is available for consultation when I have a case crisis	Supportive
Q29	Provides a safe place to talk about feeling overwhelmed	Supportive
Q42	Is available to me when I have a problem	Supportive
Q14	Takes an interest in me as a person	Supportive
Q20	Pitches in and helps handle emergencies	Supportive

Note. Continued next page

Table 9 (continued)

Retained Items for Experience of Supervision Scale and Expert Panel Categories

	Revised scale items	Expert panel category
Factor 2 (Emotional support/ Availability) (continued)		
Q30	Facilitates good team work	Supportive
Q50	Is accepting when I am not able to get everything done on time	Supportive
Q3	Models appropriate personal/ professional boundaries	Educational
Q45	Supports taking time off to deal with family emergencies	Supportive
Factor 3 (Administrative activities)		
Q27	Monitors progress towards deadlines	Administrative
Q55	Holds me accountable for completing my work on time	Administrative
Q12	Critiques my documentation to improve quality and completeness	Educational
Q28	Consults with specialists or attorney when clarification regarding policy is needed	Administrative
Q41	Filters policy and practice changes so I get exactly the information I need to do my job	Administrative
Q1	Works strategically to improve efficiency within the agency	Administrative
Q48	Strengthens collaborative relationships with community partner agencies	Administrative

factor, nine of the 10 items were previously categorized as supportive functions. For the third factor, six of the seven items were previously categorized as administrative functions.

However, for the first factor, although six of the 15 items were previously categorized as educational functions, six were categorized as supportive and three were categorized as administrative. Closer review of the items suggests that the factor is capturing something beyond the educational category described in the casework model of supervision. The majority of the items which loaded strongly on the first factor are supervisory behaviors which facilitate improved practice or improved job performance on the part of the supervisee. This includes educational functions. But administrative efforts to manage caseload size also correlate strongly.

Additionally, six items which were initially categorized as supportive, loaded strongly on this factor. These items appear conceptually different from the "supportive" items which loaded onto the Supportive factor. Those items are conceptually related around themes of supervisor availability and provision of emotional support (which are consistent with the theoretical model). The supportive items which loaded most strongly on Factor 1 are more focused on support provided by supervisors to facilitate improved practice. For example, "When assigning cases, is sensitive to the kinds of cases I prefer," the item which loaded the most strongly on this factor, can be interpreted as "assigns to me cases with which I will be successful." "Comes to court with me when I have a challenging court case" can be interpreted as "comes to court with me to make sure that everything goes well."

The three factors derived appear to be conceptually and empirically distinct. The three subscales include Performance support, Emotional support/Availability and Administrative activities. An unfortunate change in terminology is required as the researcher is unable to identify an adjective which encompasses Performance support and, therefore, in order to preserve grammatical agreement between the subscale titles. "Supportive" was renamed Emotional support/Availability. "Administrative" was renamed Administrative activities. Finally, the first factor was labeled Performance support.

Performance support describes those supervisory behaviors which actively promote good practice on the part of the child welfare workers. These behaviors include training specific activities, observing practice and teaching new skills; those activities described by Kadushin as educational. It also includes activities aimed at creating or supporting a work environment which does not overburden workers with unmanageable caseloads and is supportive of taking vacation

appropriately. Performance support includes activities like checking in with the worker about how they are feeling about their work and what they are receiving through supervision.

Emotional support/availability describes those activities described by Kadushin as supportive. These activities include being available to the worker when problems and emergencies arise. They also include providing encouragement, rather than criticism or blame when workers are struggling. These activities occur primarily at an interpersonal level and are the function of strong supervisor-supervisee relationship (e.g. My supervisor ... takes an interest in me as a person).

Finally, administrative activities are those supervisory tasks which are required to keep the program running smoothly. They include knowing and effectively applying law and policy and insuring that deadlines and protocols around documentation are adhered to. These activities facilitate efficient practice and serve to protect the agency, the staff and clients from costly mistakes. Additionally, administrative activities include "big-picture" planning, advocacy and collaborative efforts which improve efficiency and service efforts at the agency or community level (e.g. My supervisor … works strategically to improve efficiency within the agency and My supervisor … strengthens collaborative relationships with community partner agencies.)

Revised Scale

The revised version of the Experience of Supervision scale contains 32 items. In the revised scale there are 15 items in the Performance support subscale, 10 in the Emotional support/ Availability subscale, and seven in the Administrative activities subscale. The total possible score for the scale was 160, the minimum score was 32. Scores for this sample ranged from 160 to 50. The mean was 120.8 with a standard deviation of 26.2. Because the subscales each contain a different number of items, their scores are not comparable, but the scores for each

should provide meaningful data. The subscale Performance support has a maximum score of 75 and a minimum of 15. For these data, the mean was 49.9. The standard deviation was 14.8. Scores ranged from 16 to 75. The subscale Emotional support/Availability has a maximum score of 50 and a minimum score of 10. For these data, the mean was 42.2 and the standard deviation was 8.0. Scores ranged from 14 to 50. Finally, for the Administrative activities subscale the maximum value is 35, the minimum is seven. For these data, the mean was 28.65 and the standard deviation was 5.75. Scores ranged from 11 to 35.

Assessment of Construct Validity

Factorial validity. Factor analysis provides strong support for the Experience of Supervision scale possessing the three underlying functions of supervision as described by Kadushin and guiding the development of the scale. There is a meaningful three factor solution and it is generally consistent with Kadushin's model. Items were expected to load on factors which were reflective of the conceptual domain they represented. Examination of the internal structure of the scale produced two dimensions which were strongly associated with two of the dimensions of Kadushin's model: supportive and administrative. The first factor, while not entirely consistent with Kadushin's model, contains conceptual elements of Kadushin's educational supervision category. This will be discussed further in Chapter 5 in regards to the utility of the Experience of Supervision Scale. However, the factor analysis procedure strongly suggests that the Experience of Supervision scale possesses factorial validity.

Divergent validity. Multiple regression analysis yields an overall correlation coefficient which can be used to assess the degree to which the predictor variables (altogether) account for the variance in the criterion variable. The greater the R-squared value the greater the influence of the predictor (independent) variables on the criterion (dependent) variable.

Multiple regression will also weight each predictor variable. The higher the beta weight of a particular variable, the greater its relative effect on the criterion variable when all the other variables are controlled. Multiple regression permits evaluation of the relative importance of each of the predictor variables in relationship to the dependent variable.

There is no hypothesized relationship between the demographic characteristics of child welfare workers and their experience of supervision. Therefore, in order to assess divergent validity, the researcher was testing the null hypothesis, that is, that there is no relationship between the predictor variables (demographic variables) and the dependent variables (scale and subscale scores). Proving the null hypothesis is a strong indicator that the Experience of Supervision Scale is measuring something other than the differences between child welfare workers related to their demographic characteristics.

Four models were evaluated. Each contained the independent variables: gender, age range, education level, social work degree, program area and length of employment. Age range and length of employment were measured at interval level and were therefore appropriate for use. Nominal variables were dummy coded as follows: gender (male=0, female=1), education level (High school, some college, two yr degree, Bachelors=0, Masters=1), social work degree (no=0, yes=1), and program area (CPS and FC Prevention=0, FC and adoption=1) The dependent variables were the subscale scores (Performance support, Emotional support/ Availability, and Administrative Activities) and the total Experience of Supervision scale score.

The data had previously been evaluated for multivariate assumptions related to factor analysis. All cases with missing data in the items were previously eliminated. However, there were additional cases for which respondents declined to answer demographic questions (IVs). An additional six cases were eliminated in the regression analysis (n=159). Outliers were also

previously evaluated and determined not to be problematic in these data. Evaluation of homoscedasticity determined that the data contain a low amount of heteroscedasticity; the data have a slight positive skew. While homoscedasticity is assumed, regression is robust for moderate violations of homoscedasticity.

The data were evaluated for colinearity as appropriate for multiple regression. Two of the IVs, length of time employed in child welfare and age range, were correlated (r = .580) and colinearity statistic indicated both have eigenvalues close to 0 (.060 and .035). However, the condition indices are within the acceptable range for colinearity. The condition index for length of time employed is 9.310 and the index for age is 12.264. Both of these values are below 15, which is the threshold for a possible problem with colinearity, and well below the threshold of 30, where a serious problem with colinearity would be suspected.

A multiple regression analysis was performed to estimate the effects of the demographic variables on the three dimensions of the Experience of Supervision scale (performance support, emotional support/availability and administrative activities) as well as the dimension of supervision as measured by the total scale scores. The regression analyses suggest that none of the demographic variables are significantly correlated with the subscale and scale scores. The adjusted R squared values are summarized in Table 10. In none of the models is more than 6.5% of the variance explained by all of the IVs together. For the total Experience of Supervision scale score, only 4% of the variance is explained by the demographic items entered into the model. The relationship is not statistically significant ($F_{6,152}$= 2.150, p>0.05) (see Table 10).

Table 10

Multiple Regression Model Summaries for the Experience of Supervision Scale and Subscales

Scale	*R*	R^2	Adjusted R^2	Std. error of the estimate	*F*	Sig
Performance support	.317	.100	.065	14.16971	2.830	.012
Emotional support/ availability	.232	.054	.016	7.93864	1.437	.204
Administrative activities	.216	.047	.009	5.78045	1.244	.287
Total scale	.280	.078	.042	25.65571	2.150	.051

The multiple regression analysis finds the best fitting equation to describe the relationship between the independent variables and the dependent variable (regression line). The regression coefficients (*b*) are the weights provided by the multiple regression equation. A partial regression coefficient is the average amount of change that is expected to occur in the dependent variable per unit of change in the independent variable when all other independent variables' effects are controlled. The Beta weights are regression coefficients for standardized data. Standardized partial coefficients are interpreted similarly to raw coefficients, except in terms of *z* scores rather than raw scores. The Beta weights' relationship to an increase in the independent variable values are described in terms of standard deviations. Typically, raw scores are used when the units of measurement have real-world values (e.g. dollars, years) while the standardized scores are used for arbitrary units (scale scores). Beta scores are also useful in comparing the relative importance of an individual independent variable; higher scores indicate a more powerful explanatory variable in the model.

The null hypothesis is that the regression coefficient is equal to zero in the population. A statistically significant *p* value indicates that the corresponding predictor variables make significant independent contributions to the dependent variable. When the model is found to significantly explain the dependent variable scores, significant contributions from individual predictor variables further support the hypothesis that each of the predictors has some unique causal effect on the dependent variable (Grimm & Yarnold, 1995).

The first multiple regression focused on performance support. The results revealed the predictor variable Program Area makes a statistically significant contribution to the performance support score ($\beta = .244$, $p < .01$). None of the other independent variables made a statistically significant contribution. The overall R^2 for the model was .065 which was not found to be significant (see Table 11).

The second multiple regression focused on emotional support/availability. For this analysis, none of the independent variables made a significant contribution to the model. The regression analysis for administrative activities also failed to identify any significant independent variables which contributed to the model.

For the final analysis, supervision as measured by the Experience of Supervision scale was regressed on five independent demographic variables. Program area was found to make a significant independent contribution to the model ($\beta = .207$, $p < .05$). However, the overall model results ($R^2_{adj} = .042$) are consistent with the decision to fail to reject the null hypothesis that changes in the model's independent variables (i.e. demographic characteristics) do not predict changes in the model's dependent variable (Experience of Supervision scale score). This is the hoped for result, since it is desirable to have an instrument that is affected by supervisor behaviors but performs consistently across the demographic characteristics of supervisees.

Table 11

Multiple Regression Coefficients for the Experience of Supervision Scale and Subscales

	Unstandardized coefficients		Standardized coefficients			95% CI for *B*	
Scale	*B*	Std error	Beta	t	Sig.	Lower bound	Upper bound
Performance support							
Total	52.003	4.968		10.468	.000	42.188	61.817
Program area	7.128	2.310	.244	3.085	.002*	2.563	11.692
Ed level	-5.103	2.581	-.170	-1.977	.050	-10.203	-.004
SW degree	.378	2.550	.013	.148	.882	-4.661	5.416
Gender	-1.160	3.937	-.023	-.295	.769	-8.938	6.617
Length of employment	.347	1.467	.023	.236	.813	-2.551	3.244
Age range	-1.506	1.186	-.122	-1.270	.206	-3.848	.837
Emotional support/ Availability							
Total	45.221	2.783		16.248	.000	39.722	50.720
Program area	2.053	1.294	.129	1.586	.115	-.504	4.610
Ed level	-1.682	1.446	-.102	-1.163	.247	-4.539	1.175
SW degree	.224	1.429	.014	.157	.876	-2.599	3.047
Gender	-1.767	2.205	-.065	-.801	.424	-6.124	2.591
Length of employment	.244	.822	.029	.297	.767	-1.380	1.867
Age range	-1.100	.664	-.163	-1.655	.100	-2.412	.213
Administrative activities							
Total	26.970	2.027		13.308	.000	22.966	30.974
Program area	1.616	.942	.140	1.715	.088	-.246	3.478
Ed level	-.696	1.053	-.058	-.661	.509	-2.777	1.384
SW degree	-1.182	1.040	-.101	-1.136	.258	-3.237	.874
Gender	1.784	1.606	.090	1.111	.268	-1.389	4.957
Length of employment	.109	.598	.018	.181	.856	-1.073	1.291
Age range	-.098	.484	-.020	-.202	.840	-1.053	.858

Note. Continued next page. *p<.01 **p<.05

Table 11 (continued)

Multiple Regression Coefficients for the Experience of Supervision Scale and Subscales

	Unstandardized coefficients		Standardized coefficients			95% CI for *B*	
Scale	*B*	Std error	Beta	t	Sig.	Lower bound	Upper bound
Total scale score	124.194	8.995		13.807	.000	106.423	141.965
Program area	10.797	4.183	.207	2.581	.011**	2.532	19.061
Ed level	-7.481	4.673	-.139	-1.601	.112	-16.714	1.752
SW degree	-.580	4.617	-.011	-.126	.900	-9.703	8.542
Gender	-1.142	7.128	-.013	-.160	.873	-15.224	12.940
Length of employment	.699	2.655	.026	.263	.793	-4.547	5.945
Age range	-2.703	2.147	-.122	-1.259	.210	-6.945	1.538

Note. *p<.01 **p<.05

Summary of Findings

The goal of this study was to develop an instrument with psychometric properties appropriate for use in complex multivariate research regarding or including child welfare supervision. Kadushin's model of case work supervision provides a conceptual framework for understanding child welfare supervision and was used as the theoretical model for scale development. An initial administration of the draft Experience of Supervision Scale yielded a three factor solution which while not entirely consistent with Kadushin's model, has enough overlap to capture the essence of the casework supervision model. The refined version of the Experience of Supervision scale demonstrates a high level of internal consistency. Finally, construct validity was assessed through regression of subscale and total scale scores on the independent variables of gender, age range, education level, social work degree, program area and length of employment. The resulting models suggest divergent validity; the independent

variables demonstrate little relationship to the scale scores. Factorial validity also contributes to construct validity. In summary, while there is still work to be done in assessing the validity and reliability of the Experience of Supervision Scale, the current study indicates its appropriateness for use as a standardized measure.

Chapter 5: Discussion and Implications

This study proposed to develop an instrument to measure supervisory behaviors from the perspective of the child welfare worker. In order to better understand the process of child welfare supervision and how specific supervisory behaviors affect workers' behaviors and attitudes, a theoretically derived, behaviorally based measurement of child welfare supervision was developed and tested for reliability and validity. In this chapter, the limitations of the current study will be presented, and the utility of the conceptual model and implications for social work practice will be discussed. Suggestions for future research will conclude the chapter.

Limitations of the Study

Sample size. The most significant limitation of this study is the small sample size. The initial goal was to obtain at least 300 completed questionnaires. However, after several months of soliciting affirmative responses from local agency Directors, a decision was made to discontinue efforts rather than run to the risk of offending by continuing to attempt to solicit their attention. Additionally, the dates of the drawings for the gift card incentives had been previously established and once those had passed, it seemed less likely that workers contacted later in the process would be as motivated to participate. At that point, Surveymonkey.com indicated that there had been 242 respondents to the study, so it seemed appropriate to stop collecting data.

However, as noted in Chapter Four, of those 242 respondents, only 165 cases were both appropriate for the sample and contained complete data. It would have been possible to use

imputation to address missing data and increase the sample size slightly. However, a more conservative approach was chosen as this is a first attempt to establish the psychometric properties of a new instrument.

DeVellis (2003) and Worthington and Whittaker (2006) recommend a sample size of 300 for adequate factor analysis. Furthermore, Grimm and Yarnold (1995) recommend that the sample should be at least 5 times the number of variables for factor analysis, which for this case would have been 305 cases. A sample of 165 cases is well below these recommendations. On the other hand, Hinkin (1995) suggests that a sample of 150 is adequate for exploratory factor analysis.

Worthington and Whittaker (2006) provide two principles for determining whether the sample developed was adequate for the factor analysis performed which are relevant to this study: 1) sample sizes of 150 to 200 are likely to be adequate if the data set yields communalities higher than .50 or 10:1 items per factor with factor loadings of at least 0.4, and 2) smaller samples may be adequate if factor analysis yields communalities of .6 or greater and or at least 4:1 items per factor and factor loadings greater than 0.6 .

For this sample, the initial three factor PCFA with varimax rotation resulted in three factors with more than 10:1 items per factor with the majority of the factor loadings greater than .50. However 11 of the 61 items' communalities were less than .50 (see Appendix K). This suggests that the sample, while not ideally "large," approached large enough for this exploratory study. Further support for this assessment is that the KMO measure of sampling adequacy was good (> .60).

Sample characteristics. For scale development, sample size is considered more salient than representativeness. Purposive sampling is frequently used in scale development studies to

generate large samples while arranging for the characteristics of the population of interest to be included. For this study, a purposive sample of child welfare workers from across Virginia was developed. Review of the localities for which a contact person was designated to forward the invitation to participate in the study to their staff suggests that workers from large and small, urban and rural localities from all five geographical regions of the state are included in the sample. In this study, the distribution of responding agencies can be considered a proxy for the distribution of respondents in terms of agency characteristics.

The known sample characteristics suggest that the sample is reasonably diverse; that is, there is variability in terms of age, length of experience, and program area. However, because so little information is known about the population of child welfare workers in the state of Virginia, although some demographic variables were collected, it is not possible to compare the sample to the population on more than a few characteristics. A significant limitation of this study is a lack of information regarding the sample's characteristics in comparison to the larger population.

In addition, because the sampling method did not permit tracking of which invitees responded compared to which invitees declined to respond, it is not possible to determine if there are significant differences between the two groups. One potential factor influencing who responded was the use of the local contact person to forward the invitation. While this sampling method was designed to protect the anonymity of the invitees and respondents, it required that an intermediary who was known to the study invitees forwarded the email invitation. In many cases, the agency director identified a child welfare supervisor as the local agency contact. The workers therefore received the invitation to participate in the study through a supervisor, maybe their own. Receiving the invitation from a known person may have influenced the degree to which invitees felt confident that their responses would be confidential. The workers' feelings

about the person from whom they received the invitation may also have influenced whether they chose to respond or not.

Measurement error. Measurement always contains some element of error. When measuring a construct, the goal is to use an instrument that has good reliability and validity and thereby minimize the degree to which measurement error will affect results. When developing an instrument to measure a construct, it is important to acknowledge potential sources of error and to assess how well the measure appears to have kept errors to a minimum. Chance error or random error can undermine the reliability of an instrument, while systematic error can compromise validity.

Reliability is the degree to which an instrument consistently measures the same thing. Random error can include respondents marking the wrong selection, coding errors, ambiguous instructions, and so forth (Rubin & Babbie, 2001). This type of error is not systematic; it is as likely to have the effect of overestimating as underestimating a particular value. The concern with random error is that it may introduce variance into the data that is not a reflection of the true variation across individual responses (Carmines & Zeller, 1979). For this study, the Experience of Supervision scale produced high alpha coefficients. These measures of internal consistency are inversely related to random error and therefore suggest that random error is not a significant concern in these data. The high alpha coefficients also suggest that the scale is likely to perform reliably in additional administrations.

Nonrandom or systematic error creates a biasing effect on measurement because the measurement errors always occur in the same direction. Nonrandom error poses a significant threat to validity; the instrument is prevented from measuring the construct it is intended to

measure. Instead, the data may represent something else entirely. Nonrandom errors of concern in scale development particularly include those related to response set errors.

Response set errors refer to those patterns of responses which occur irrespective of the content of the items. The bias resulting from a respondent answering all the questions in the same way (either agreeing or disagreeing) regardless of the question content is called the acquiescent response set (Rubin & Babbie, 2001). The tendency to agree or disagree with items regardless of how the respondent actually feels about the items will result in extreme scores and the potential for distortions of estimates and means. One way in which this bias can be reduced is through use of positively and negatively worded items in the scale. If there are an equal number of positively and negatively worded items, the acquiescent respondents will tend to have middle scores which will have much less of a distorting effect on the data. Additionally, items containing similar ideas worded both positively and negatively can be compared in order to identify potential cases of acquiescent response set (Spector, 1992).

For this study, only positively worded items were used. Because this was a first attempt to develop the Experience of Supervision Scale there were more items than ideal for a single standardized instrument already. Long questionnaires contribute to participant drop out. Making the questions more complex would have added to the length of time the questionnaire took to complete, perhaps increasing drop out rates as well. Additionally, negatively worded items can be "bad" items; they are easy to misread or to misunderstand, producing hard to interpret results. In this first attempt at scale validation, it seemed more important to assess how the items functioned together as a measure of the construct, than to introduce another element of potential error. However, in further administrations, it would be ideal to reword some items to reverse the

scoring or to add inversely worded items containing similar ideas to permit the research to assess for acquiescent response set.

Social desirability bias refers to a pattern of responses which are not reflective of the respondent's true belief, but instead, of positions that the respondent believes will make him or her look good in the researcher's eyes. This form of systematic bias is particularly problematic in face-to-face interviews and when the subject matter has potential to be embarrassing. The researcher can address concerns about social desirability bias through thoughtfulness in the wording of questions as well as emphasizing confidentiality or anonymity through the sampling method employed. The researcher can also add a social desirability scale to the questionnaire containing the new scale's items (Rubin & Babbie, 2001). Data can be assessed to determine if there are items which strongly correlate with the social desirability scale; these items can be considered for deletion.

For purposes of this study, a social desirability scale was not used. As previously stated, there were concerns about scale length and complexity. Further, the use of a second scale introduces the possibility that the respondent's answers to the new scale's items may be influenced by the experience of answering the second scale's questions. Additionally, a social desirability scale was not considered crucial by the researcher because careful attention had been given to protecting the anonymity of respondents through the sampling procedure. Furthermore, the scale items were not particularly personal or embarrassing. The risk of participants' responses being influenced by social desirability bias was assessed as being particularly low.

However, in future studies if respondents were not anonymous, social desirability bias may be more of concern. Workers might be less accurate in their responses regarding their supervisors' behavior if they believe they would be perceived as poor employees by the

researcher. In another administration it might be possible to include a social desirability scale for half of the respondents, enabling the researcher to assess the degree to which social desirability may be an issue with these data and this population, and also the degree to which the inclusion of a second scale may influence the results of the Experience of Supervision scale. As with all research concerning supervision, careful consideration would have to be given to issues regarding confidentiality and privacy; if workers believe that there could be reprisals for indicating that their supervisor has deficits or areas where behaviors are not as frequent or well-developed as other supervisors, responses are less likely to be truthful.

Finally, the validity of the Experience of Supervision scale was primarily assessed through factor analysis. The derived three factor model which replicates Kadushin's three function model of casework supervision lends strong support to the scale's having validity. Additionally, evaluation of the relationship between the demographic variables and the scale and subscale scores suggests that the scale performs consistently across demographic variables; scale scores are affected by supervisory behaviors rather than the characteristics of the respondents. This was the desired result of this initial evaluation of validity, however additional work should be done.

In particular, in future administrations it would be appropriate to assess the construct validity of the Experience of Supervision scale further. The scale could be administered with measures of satisfaction with supervision, for example, to assess convergent validity. The efforts reported in this study are reasonable first attempts to establish validity, but further efforts to assess validity could strengthen the claim that the scale has validity.

Utility of the Conceptual Model

Summary of conceptual issues. The goal of this project was to develop a scale with psychometric properties appropriate for use in complex, multivariate research that measured supervisory behaviors as experienced by child welfare workers. In order to accomplish this, a model of social work supervision was selected to serve as the conceptual framework for item generation and selection. The project is considered a success in that the factor structure extracted is consistent with the underlying theoretical model.

Defining the construct to be measured by a new scale is probably the most important step of the scale development process. Without careful and specific delineation of the construct to be measured prior to item generation it is impossible to insure that items generated are indicators for that construct and no other. It is, further, impossible to assess the ultimate validity of a scale if there is no conceptual definition of the construct to compare it to.

In this case, the casework supervision model developed by Kadushin was selected as the conceptual framework. It was selected for this study because in addition to its relevance as a social work supervision model, it was initially developed out of data Kadushin collected in interviews with social workers in practice in public agencies, including child welfare agencies, about their experiences of supervision. The model categorizes supervisory behaviors into three functions; a good supervisor has well-developed skills and regularly performs activities in all three categories. Although the definitions of the categories have been provided previously, they will be reviewed here briefly to facilitate comparison with the results of this study.

Administrative functions. In the child welfare field, administrative supervision includes monitoring the casework of child welfare workers to insure that time lines and financial procedures are consistently followed. Administrative supervision is generally focused on agency

policy and public accountability; supervisors coordinate, direct, and monitor social work practice. Other administrative functions include managing caseloads/case assignments, administering performance evaluations, and addressing performance problems. Additional administrative functions include middle-management activities such as understanding, supporting, and communicating the goals of the larger agency to the child welfare staff and also identifying and advocating with upper management for strategic efforts to improve systemic and individual worker effectiveness (Kadushin & Harkness, 2002).

Supportive supervision. Kadushin and Harkness (2002) define supportive supervision as "concerned with helping the supervisee deal with job-related stress and developing attitudes and feelings conducive to the best job performance" (p. 277). Job stress is generated by work with difficult clients, ambiguous or conflicting roles and responsibilities, constant change, and unreasonable personal expectations about job performance. Good supportive supervision effectively mediates between the structural stresses of the work environment and the supervisees (Kadushin & Harkness, 2002; Smith, 2005; US GOA, 2003). In providing supportive supervision, the supervisor is available to the worker for the ventilation of feelings, to provide positive feedback, and to hold out reasonable expectations in order to reduce, ameliorate, and prevent stress. Additionally, the supervisor can take an active role in facilitating positive peer interaction through leading regular staff meetings and/or group supervision, developing peer supervision and consultation opportunities, and structuring mentoring relationships (Kadushin & Harkness, 2002; Wehrman et al., 2002).

Finally, the supervisor can actively engage in the process of trying to improve the organizational context through facilitation of better working relationships across teams or better interagency relationships, for example. The supervisor can advocate for reduced caseloads,

recognition for outstanding performance (Kadushin & Harkness, 2002), and policies which support a specific multi-level response to crises for the staff (e.g. death of a client, or worker being threatened of assaulted) (Gustavsson & MacEachron, 2004).

***Educational functions*.** The chief function of educational supervision is the enhancement of the supervisee's practice skills through direct provision of knowledge and skills or by making training available and supporting gains made. The supervisor facilitates the application of new knowledge and skills to ongoing case practice through case consultation. Additionally, the supervisor engages in continual assessment of each supervisee's current skill level, works with the supervisee to develop an individualized learning plan which will facilitate improved practice and professional development (Kadushin & Harkness, 2002), and frames training as an "opportunity", further increasing the likelihood of training transfer (Salas & Cannon-Bowers, 2001; Wehrmann et al., 2002). The supervisor can also play an active role in advocating with administration for increased training funds and agency policy and practices which support educational goals, innovation, and worker self-efficacy (Kadushin & Harkness, 2002).

Factor solution fit with conceptual model. The second and third factors extracted are a very good fit with the conceptual model's categories of administrative functions and supportive functions. The items which loaded onto the second factor, Emotional support/Availability, were predominantly items which had previously been identified as supportive by the researcher and by the expert panel. They included items addressing ventilation of feelings and encouragement of self-care. Additional items addressed the supervisor's availability both physically (My supervisor … pitches in when there is an emergency) and emotionally (My supervisor … is available to me when I have a problem).

The supportive items which loaded on this factor tend to be those functions which are provided through a strong, positive supervisor-supervisee relationship. One item, "models appropriate personal/professional boundaries," was sorted by the researcher and expert panel into the educational category. Its inclusion in the Emotional support/Availability category in this study suggests that a supervisor with good boundaries is better able to provide emotional support. The supervisor's modeling of good boundaries is not experienced as a learning opportunity as much as the characteristic lends itself to the worker experiencing emotional support. This might be an important area to explore with future research.

The item "facilitates good teamwork" which also sorted onto the second factor, is consistent with the nominal definition for the category for supportive functions in Kadushin's model. However, it is the only item retained for the Emotional support/Availability subscale which addresses the supervisor's role in the development of supportive practices and policies in addition to the provision of emotional support.

Items selected into the Administrative activities subscale include items addressing the monitoring of documentation and timeliness of work activities. Other items address the supervisor's role in imparting guidance to the worker regarding policy and practice guidelines. Finally, several items address strategic planning and community level interventions (strengthening collaborative relationships with partner agencies).

The item "critiques my documentation to improve quality and completeness" was initially sorted by the researcher into the Administrative functions category based on the nominal definition provided by the model. Oversight of documentation is traditionally an administrative activity. However, the expert panel consistently sorted this item into the Educational functions category based, perhaps, on the implied training opportunity in the term "critiques." In this

study, the item loaded on the Administrative activities subscale, suggesting that workers experience this function similarly to other administrative functions, rather than as an opportunity for skill building. In any case, the seven items in the revised Experience of Supervision scale which make up the Administrative activities subscale are consistent with the model's administrative category of supervisory behaviors.

The first factor extracted, Performance support, is generally consistent with Kadushin's casework supervision model in that it includes six items which were previously sorted into the educational category and are clearly related to training and/or the supervisor's efforts to enhance the skills of the worker. These behaviors include suggesting trainings to attend, observing the workers' behavior in the field in order to encourage skill development, and using role play to teach new skills. Additionally, after revision, this factor contains the greatest number of items previously categorized as educational. However, items previously classified as both administrative and supportive loaded more strongly than educational items onto this factor. Additionally, when combined the six supportive items and the three administrative items outnumber the six educational items.

The three administrative items which sorted onto this factor are similar to each other in their focus on supervisory behaviors which promote good practice on the part of the worker. They include efforts to insure that caseloads are manageable and to address resource gaps in the community. They are systems interventions intended to improve workers' abilities to do their jobs well. These items loaded particularly strongly on this factor, and suggest that the construct which underlies this factor encompasses more than just training or skills development.

Six items which were previously categorized as supportive also loaded strongly onto the Performance support factor. These items appear to differ from the items sorted onto the

Emotional support/Availability in their focus on the supervisor's efforts to enhance workers' performance rather than attend to workers' feelings. As noted above, the casework supervision model includes the development of supportive practices and policies as a supportive function. In this sample, several of these activities sorted into the Performance support subscale.

For example, the item which loaded most strongly onto the first factor, "when assigning cases, is sensitive to the kinds I prefer" was categorized by the expert panel as supportive, although the researcher had categorized it as administrative. It was intended to address the supervisor's administrative role in insuring that cases are handled competently by making case assignments to workers with the requisite skills to manage the case. However, assigning cases to workers who possess the skills to manage them effectively also increases the likelihood that the workers will have positive experiences with those cases. This is perhaps the reason the expert panel categorized this item as supportive. When looking at this item in relationship to the other items loaded onto the first factor, it appears that the workers experience this behavior as supportive of their doing good work.

The item "comes to court to support me when I have a challenging case" was intended to capture the supervisor's efforts to be available to the worker for emotional support, if needed, relative to a challenging court proceeding. However, when looking at this item in relationship to the other items retained on this factor, there is a suggestion that the workers experience this supervisory behavior either as an opportunity to learn something through observing their supervisor's behavior in court or as an assurance that the supervisor will step in if the worker needs practical assistance. Both interpretations suggest that the behavior is an intervention into a challenging situation which the worker experiences as an aid to doing their job.

The item that seemed to make the least sense conceptually in this subscale was "encourages me to take vacation." This item was previously categorized as supportive and at face value seems to have more in common with the other items that sorted onto the Emotional support/Availability subscale. However, one of the chronic stressors in the field of child welfare is inadequate staffing. It can be a challenge to relieve a worker of all responsibilities for an extended period of time; it is not uncommon for workers to be expected to be available by phone even when out of the office. Further, they may feel responsible to put in extra hours before and after a vacation to insure that all their responsibilities are covered adequately. A supervisor who is willing to cover a worker's caseload or who actively works to develop a coverage plan may be experienced as promoting the worker's ability to actually experience vacation as a break from work. This item's presence on the Performance support subscale suggests that being able to take vacation enhances workers' abilities to do their jobs; a supervisor's being encouraging of taking a vacation may include the institution of local policies or procedures which insure adequate coverage for cases while workers are on vacation.

Revised conceptual model. It is not possible through a single administration of the Experience of Supervision scale to draw definitive conclusions about the integrity of the underlying conceptual model. However, the findings of this study provide strong support for the scale's ability to assess aspects of the casework supervision model. The factor analysis results indicate that the scale contains three dimensions which encompass the three supervisory functions conceptualized by Kadushin: administrative, supportive, and educational.

The findings further suggest that these three dimensions include some conceptual differences from the model's. These differences are seen as potentially adding value to the scale, as their distinctions may provide valuable information through future research. Supportive

supervision is frequently cited in the literature as a variable which influences job satisfaction, worker morale, and retention. Further, the absence of supervisory support has been identified as a leading contributor to burnout and worker turnover. The factor analysis indicates a distinction between Emotional support/Availability and Performance support behaviors as experienced by the workers. This distinction has potential to provide essential information about which "supportive" behaviors are most closely correlated to a variety of positive and negative outcomes for workers.

Additionally, the ultimate goal of improving the quality and effectiveness of the child welfare workforce is to improve outcomes for children and families served by the system. Supervisory activities are believed to be instrumental in reducing turnover, but also in enhancing the practice skills of their supervisees. The Performance support subscale operationalizes the idea that supervisory behaviors which actively support good practice go beyond training and transfer of training activities. This subscale has the potential to permit careful evaluation of the relationship between supervisory behaviors which support enhanced worker performance and improved client outcomes.

Implications for Social Work Practice

The field of child welfare is uniquely associated with the social work profession. It, more than any other practice area, has been developed and staffed primarily by social workers since its infancy (Lindsey, 2004; Schorr, 2000). Early child welfare workers were predominately social workers by training. These child welfare social workers were generally well educated and they were allied with the mental health and medical professionals of their day. In fact, the earliest child welfare workers were, of course, pioneers in the field of social work (Woods & Hollis, 2000). However, in the late 60s and 70s, in response to increased public focus on child abuse and

mandated reporting laws, the number of child abuse reports made to public agencies escalated dramatically. As the child welfare system subsequently became more risk-averse and liability-minded, the numbers of children entering foster care soared. Equally problematic, the child welfare system, already under-resourced, was forced to reduce efforts directed towards working with families to keep children in their homes, and to rely increasingly on less-skilled staff to keep up with the volume of child protection work (Lindsey, 2004; Schorr, 2000; Zlotnik, 2002). Intense public scrutiny coupled with inadequate resources resulted in a decline in the public's positive regard for the child welfare system and the workers who staffed it (Zlotnik, 2002).

In the years that followed, there were numerous attempts to improve the child welfare system through legislation; laws and policies were enacted that directed the day-to-day activities of the child welfare workers, reducing professional autonomy and empowerment, and increasing judicial oversight. The Child Welfare and Adoption Assistance Act of 1980, established expectations for the public agencies regarding the removal and placements of children, but did not provide the resources necessary to meet these expectations. As a result, the reputation of the child welfare system was further tarnished and an exodus of professionally trained social workers from the field began.

At the same time, the public's expectations remained high and child welfare "failures" resulted in intense media coverage, class action lawsuits, and eventually court action which resulted in some state child welfare systems being placed in receivership or under the direct supervision of a court (US GAO, 2003; Zlotnik, 2002). The Adoption and Safe Families Act of 1997 specified key performance goals designed specifically to insure child safety and to reduce the time within which children are placed in safe and permanent homes. This legislation further raised the stakes for public child welfare agencies; states face fiscal penalties for failing to meet

performance guidelines and demonstrate improved outcomes for children and families (Child Welfare Information Gateway, n.d.).

This study is a small contribution among many contributions to a much bigger goal: to improve the outcomes for the children and families served by the child welfare system. In 2006(a), the Children's Defense Fund published *Components of an Effective Child Welfare Workforce to Improve Outcomes for Children and Families*. In this guide, 14 critical components are identified:

1) Child welfare agencies must be led by strong, competent, visionary, and committed child welfare professionals.
2) A supportive organizational environment is needed to promote workers' long term commitment to child welfare.
3) Child welfare caseloads and workloads must be kept at manageable levels.
4) Meaningful supervision and mentoring for caseworkers enhance best practice interventions with children and families, which in turn increase caseworkers' job satisfaction.
5) The quality of the workforce must be strengthened through opportunities for quality and tailored professional preparation.
6) Specialized, competency based training and professional development are critical to the development and maintenance of a skilled child welfare workforce.
7) Timely, accurate, and consistent data are essential for supporting a quality, effective child welfare workforce.
8) Relevant, comprehensive research and evaluation are critical to ensuring continuous improvements in the child welfare workforce.

9) Effective quality assurance and accountability mechanisms must be in place to support an effective child welfare workforce to ensure positive outcomes for children, youth, and families.
10) Workers must be provided with technological resources and support staff to help them meet their professional responsibilities and keep track of important information about the children and families they serve.
11) The child welfare workforce must have a safe and suitable workspace.
12) Valuable employment incentive, including pay increase, benefits, and promotional opportunities are essential for the development of an effective child welfare workforce.
13) Agencies must become culturally competent to effectively serve and support children and families.
14) Agencies must build connections with organizations and community groups to strengthen their public image, improve outcomes, and attract professionals to the field of child welfare.

The Children's Defense Fund developed this list to point out how federal policy changes can and should support these goals. Clearly, state and local governments would also have a role in ultimately achieving these objectives. The list is included here not to suggest that somehow all of these components can be addressed by individual child welfare supervisors, but to emphasize the important role that supervisors might have in addressing these areas in the local agencies where they work. This lists further serves as a reminder about the goal of this project: to develop a tool to aid in better understanding the relationship between various supervisory behaviors and

critical workforce issues and in guiding interventions which better prepare or better support supervisors to fulfill the responsibilities of their position.

This study's contribution is unique in that it fills a gap between the theoretical literature concerning supervision and descriptive studies of child welfare workers' perceptions related to their experiences of supervision. Its use will permit assessment of the impact of different models of supervision, different practice models, and supervisor to supervisee ratios on workers' experience of supervision.

Recommendations for Future Research

Scale development is a many-staged process; cumulative evidence derived from multiple administrations contributes to better understanding of the scale's relationship to the construct that it seeks to measure. This project is a first effort to develop a scale that is appropriate for use in complex, multi-variable research. The findings described here yielded a 32 item instrument with 3 distinct subscales. The factor structure is appropriately parsimonious and guided scale refinement, especially reduction of items. The revised scale and subscales are consistent with the conceptual model upon which the items were based. Additionally, the items together demonstrate reliability well above the acceptable range and initial estimates of validity suggest the scale has acceptable validity.

Ideally, with additional administrations further evidence of reliability and validity would emerge. In particular, the scale would benefit from a more explicit assessment of construct validity; the correlation between the scale and subscale scores with measures of job satisfaction or satisfaction with supervision would be salient. The suggestion of Worthington and Whittaker (2006) that items be deleted if there is less than 0.15 difference in value between the two highest factor loadings is only one strategy for handling item deletion/retention using factor analysis

results. This study's sample was also smaller than ideal. Consideration could be given to including some of the items deleted here in an administration with a larger sample, to determine if the results in this study were replicated, or if those items sorted more distinctly for the new, larger sample.

Additionally, a 5 point Likert scale was selected for this administration. In a second iteration of the scale development process, two forms of the scale could be tested: a 5 point and a 7 point Likert format. There were no concerns about the data generated in this study using the 5 point response set, but ideally the Experience of Supervision scale would be sensitive enough to measure changes in supervisors' behaviors over a period of time and/or following an intervention. A 7 point Likert scale might increase the variability of the scores, thereby increasing the scale's sensitivity. The scale's sensitivity might also be assessed through its use in pre- and post-testing for a supervision training program. If, instead, multiple forms of the instrument were to be evaluated, it might also be possible to incorporate positively and negatively worded items into one version of the scale to assess psychometric properties when addressing acquiescent response set bias.

Finally, it would be ideal to conduct a confirmatory factor analysis at some point. The result of the current study and the structure of the current Experience of Supervision scale could be utilized to specify the expected factor structure and item loadings. A confirmatory factor analysis procedure which replicated the finding of this study would provide further evidence of the Experience of Supervision scale's validity.

As has been discussed throughout this dissertation, child welfare supervision has been found to be a key variable in understanding child welfare workforce issues including retention, job satisfaction, and morale (See Dickinson & Perry, 2002; Jacquet et al., 2007; Kim, 2011;

Landsman, 2001; Mor Barak et al., 2001; Mor Barak et al., 2006; Renner et al., 2009; Smith, 2005; Travis & Mor Barak, 2010; Zlotnik et al., 2005). Supervisors have been found to be a contributing variable in organizational culture and climate, which in turn have been found to influence workers' job satisfaction and client outcomes (see Ellett, Ellett & Rugutt, 2003; Glisson & Hemmelgarn, 1998). The Experience of Supervision Scale was intentionally designed to focus solely on supervisory behaviors. Although there is much to be gained by better understanding the role of organizational culture and climate on a variety of child welfare workforce and client outcomes, this researcher believes that supervision merits evaluation as a unique variable. A standardized measure of child welfare supervision from the perspective of the workers provides an opportunity for future research to clarify our understanding about the role of supervision relative to a variety of other variables and issues as well as to facilitate the meaningful comparison of study results.

Additionally, this study suggests that the three constructs which comprise the experience of supervision from the perspective of the worker, while consistent with Kadushin's casework supervision model, are distinguished in several ways which may contribute valuable insight to these areas of research. Particularly, the relationship between client outcomes and the subscale Performance support may yield valuable information about supervisory activities which enhance worker performance. Also, the distinction between activities which are supportive of work performance versus those which offer emotional support may permit future projects to answer important questions about the relationship between both types of supportive supervisory behaviors and other variables of interest (e.g. client outcomes and retention rates).

Finally, although the goal of this project was to develop an instrument to measure child welfare workers' experience of supervision, the resulting product may actually have uses in

practice areas in addition to child welfare. The Experience of Supervision Scale ultimately contains items which pertain to child welfare supervision, without including items which are exclusive to child welfare supervision. The scale may be useful in any setting where workers are engaged in casework: nursing homes, hospitals, community mental health centers, etc. Testing the scale for appropriateness for use with other populations is another area where future research would be necessary and potentially exciting.

Conclusion

The lack of standardized measures, in general, has been noted as a problem in the area of research into child welfare workforce issues (Zlotnik et al., 2005). The rationale for this project included the following assessments: 1) that child welfare supervision from the perspective of the child welfare worker is an increasingly important variable in the study of child welfare workforce issues 2) that the lack of a standardized instrument to measure child welfare supervision from the workers' perspective is a significant limitation to developing understanding of these issues and 3) that more effectively assessing the impact of supervisory training curricula on supervisory behaviors would greatly assist the states in refining their efforts to improve the quality of supervision provided to child welfare workers.

The scale development process detailed here was exploratory in nature, but represents a significant first step towards the development of a tool which can be used in future research with the child welfare worker population. The results demonstrate that the Experience of Supervision scale is appropriate for use in its current state, although additional steps in the scale development process could be conducted.

As the states move forward in addressing the areas in which they are still not in substantial conformity with the federal guidelines assessed by the CFSRs, the stakes continue to

rise. A well-trained, stable, and effective workforce is a critical component in achieving eventual success and child welfare supervisors have a vital role to play in the development and maintenance of such a workforce. The Experience of Supervision scale can be used to develop a richer understanding of the relationship between supervisory behaviors and workforce and client outcomes. More detailed information about effective supervision could guide choices about supervisory training programs and identify where to focus national intervention efforts intended to improve worker retention and the quality of services provided and, in turn, improve outcomes for children and families.

References

Albers, E. C., Reilly, T., & Rittner, B. (1993). Children in foster care: Possible factors affecting permanency planning. *Child and Adolescent Social Work Journal, 10*(4), 329-341.

American Public Human Services Association (2005). *Report from the 2004 child welfare workforce survey: State agency findings.* Washington, DC: Author.

Balfour, D. L., & Neff, D. M. (1993). Predicting and managing turnover in human service agencies: A case study of an organization in crisis. *Public Personnel Management, 22*(3), 473-487.

Bardo, J. W., & Yeager, S. J. (1982). Consistency of response style across types of response formats. *Perceptual and Motor Skills, 55*, 307-310. doi:10.2466/PMS.55.5.307-310

Bardo, J. W., Yeager, S. J., & Klingsporin, M. J. (1982). Preliminary assessment of format-specific central tendency and leniency error in summated rating scales. *Perceptual and Motor Skills, 54*, 227-234. doi:10.2466/PMS.54.1.227-234

Bednar, S. (2003). Elements of satisfying organizational climates in child welfare. *Families in Society, 84*(1), 7-13. doi:10.1606/1044-3894.70

Bell, H., Kulokarni, S., & Dalton, L. (2003). Organizational prevention of vicarious trauma. *Families in Society, 84*(4), 463-470. doi:10.1606/1044-3894.131

Bennett, J., Lehman, W., & Forst, J. (1999). Change, transfer climate, and customer orientation. *Group and Organization Management, 24*, 2, 188-207.

Black, R. (2003). *The social work dictionary* (5th ed.). Washington, DC: National Association of Social Workers Press.

Bland, J. M., & Altman, D. G. (1997, February 22). Statistics notes: Cronbach's alpha. *British Medical Journal, 314*, 572.

Bride, B. E., Jones, J. L., & MacMaster, S. A. (2007). Correlates of secondary traumatic stress in child protective services workers. *Journal of Evidence-Based Social Work, 4* (3/4), 69-80.

Caplan, R. D., Cobb, S., & French, J. (1975). Relationship of cessation of smoking with job stress, personality, and social support. *Journal of Applied Psychology, 60*(2), 211-219.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment* [Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-017]. Newbury Park, CA: Sage.

Cearley, S. (2004). The power of supervision in child welfare services. *Child & Youth Care Forum, 33*(5), 313-327. doi:10.1023/B:CCAR.0000043038.97882.2f

Children's Defense Fund and Children's Rights, Inc. (2006a). *Components of an effective child welfare workforce to improve outcomes for children and families: What does research tell us?* Retrieved from http://www.childrensdefense.org/child-research-data-publications/data/components-of-an-effective-child-welfare-workforce.pdf

Children's Defense Fund and Children's Rights, Inc. (2006b). *Supporting and improving the child welfare workforce: A review of Program Improvement Plans (PIPs) and recommendations for strengthening the child and family service reviews.* Retrieved from http://www.childrensdefense.org/child-research-data-publications/data/supporting-and-improving-the-child-welfare-workforce.pdf

Child Welfare Information Gateway, (n.d.). *Adoption and safe families act of 1997, P.L. 105-89, Overview.* Retrieved from http://www.childwelfare.gov/systemwide/laws_policies/federal/index.cfm?event=federalLegislation.viewLegis&id=4

Child Welfare League of America. (2001). *The child welfare workforce challenge: Results from a preliminary study.* Washington, DC: Author.

Child Welfare League of America. (1995). *Standards of excellence for family foster care services*. Washington, DC: Author.

Collins-Camargo, C. (2006). Clinical supervision in public child welfare: Themes from findings of a multisite study. *Professional Development: The International Journal of Continuing Social Work, 9*(2/3), 100-110.

Collins-Camargo, C., & Groeber, (2003). Adventures in partnership: Using learning laboratories to enhance frontline supervision in child welfare. *Professional Development: The International Journal of Continuing Social Work, 6*(1/2), 17-31.

Couper, M. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly, 64*, 464-494. Retrieved from http://www.jstor.org/stable/3078739

Couper, M., Traugott, M. & Lamias, M. (2001). Web survey design and administration. *Public Opinion Quarterly, 65*, 230-253. Retrieved from http://www.jstor.org

Cyphers, G. (2003). *Report from the 2004 child welfare workforce study*. Washington, DC: American Public Human Services Association.

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? *International Journal of Market Research, 50*(1), 61-77.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.) [Applied Social Research Methods Series, Volume 26]. Thousand Oaks, CA: Sage.

Dickinson, N. S., & Perry, R. E. (2002). Factors influencing the retention of special educated public child welfare workers. In K. Briar-Lawson & J. L. Zlotnik (Eds.), *Evaluation research in child welfare: Improving outcomes through university-public agency partnerships* (pp. 89-104). Philadelphia, PA: Haworth Press.

Drake, B., & Yadama, G. N. (1996). A structural equation model of burnout and job exit among child protective services workers. *Social Work Research, 20*(3), 179-187. Retrieved from http://www.cinahl.com/cgibin/refsvc?jid=958&accno= 1997008057

Dvorak, P. (2008, February 2). New cases strain child welfare after death of four D.C. girls, firings. *The Washington Post*. Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2008/02/01/AR2008020102010.html

Erera, I. P., & Lazar, A. (1994a). The administrative and educational functions in supervision: Indications of incompatibility. *The Clinical Supervisor, 12*(2), 39-56.

Erera, I. P., & Lazar, A. (1994b). Operating Kadushin's model of social work supervision. *Journal of Social Science Research, 18*(3/4), 109-122.

Ellett, A. J., Ellett, C.D., & Rugutt, J. K. (2003). *A study of the personal and organizational factors contributing to employee retention and turnover in child welfare in Georgia.* Athens, GA: University of Georgia.

Ellett, A. J., Ellis, J., Westbrook, T. & Dews, D. (2006). A qualitative study of 369 child welfare professionals' perspectives about factor contributing to employee retention and turnover. *Children and Youth Services Review, 29*(2007), 264-281. doi:10.1016/j.childyouth.2006.07.005

Ellett, A. J., & Leighninger, L. (2007). What happened? An historical analysis of the de-professionalization of child welfare with implications for policy and practice. *Journal of Public Child Welfare, 1*(1), 3-33. doi:10.1300/J479v01n01_02

Faller, K. (2004). *Training for child welfare supervisors: Final project report*. Ann Arbor, MI: University of Michigan School of Social Work.

Fox, S., Miller, V. & Barbee, A. (2003). Finding and keeping child welfare workers: Effective use of training and professional development. *Journal of Human Behavior in the Social Environment, 7*(1/2), 67-81.

Galesic, M., Tourangeau, R., Couper, M., & Conrad, F. G. (2008). Eye tracking data: New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly, 72*(5), 892-913. doi:10.1093/poq/nfn059

Glisson, C., & Hemmelgarn, A. (1998). The effects of organizational climate and interorganizational coordination on the quality and outcomes of children's service systems. *Child Abuse & Neglect, 22*(5), 401-421. doi:10.1016/S0145-2134(98)00005-2

Griffin, C., & Shiell, D. (2003). Improving clinical supervisory practice in front line CPS supervisors: A mentoring model. *Professional Development: The International Journal of Continuing Social Work Education, 6*(1/2), 52-56.

Grimm, L. G., & Yarnold, P. R. (1995). *Reading and understanding multivariate statistics.* Washington, DC: American Psychological Association.

Gustavsson, N., & MacEachon, A. E. (2004). When a child welfare client dies: An agency-centered perspective. *Child Welfare, 83*(4), 317-40. Retrieved from www.cinahl.com/cgi-bin/refsvc?jid=699&accno=2005093417

H.R. 867, 105th Congress (1997). *Adoption and Safe Families Act of 1997*. Retrieved from http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=105_cong_bills&docid=f:h867enr.txt.pdf

Hardesty, D. M., & Bearden, W. O. (2003). The use of expert judges in scale development: Implications for improving face validity measures of unobservable constructs. *Journal of Business Research, 57*, 98-107. doi:10.1016/S0148-2963(01)00295-8

Hinkin, T. (1995). A review of scale development practices in the study of organizations. *Journal of Management, 21*(5), 967-988.

Jacquet, S. E., Clark, S. J., Morazes, J. L., & Withers, R. (2007). The role of supervision in the retention of public child welfare workers. *Journal of Public Child Welfare, 1*(3), 27-54. doi:10.1300/J479v01n03_03

Jones, J. L., Alexander, M. D., & Patterson, N. (2003). The Tennessee child protective services supervisors development project. *Professional Development: The International Journal of Continuing Social Work Education, 6*(1/2), 74-78.

Jones, L. (2002). A follow-up of a Title IV-E program's graduates' retention rates in a public child welfare agency. In K. Briar-Lawson & J. L. Zlotnik (Eds.), *Evaluation research in child welfare: Improving outcomes through university-public agency partnerships* (pp. 39-52). Philadelphia, PA: Haworth Press.

Jones, L., & Okamura, A. (2000). Reprofessionalizing child welfare services: An evaluation of a Title IV-E Training Program. *Research on Social Work Practice, 10*(5), 607-621. Retrieved from http://rsw.sagepub.com/content/10/5/607

Kadushin, A. (1974). Supervisor-supervisee: A survey. *Social Work, 19*(3), 288-298.

Kadushin, A. (1992). Social work supervision: An updated survey. *The Clinical Supervisor, 10*(2), 9-27.

Kadushin, A., & Harkness, D. (2002). *Supervision in social work* (4th ed.). New York, NY: Columbia University Press.

Kim, H. (2011). Job conditions, unmet expectations, and burnout in public child welfare workers: How different from other social workers? *Children and Youth Services Review, 33*, 358-367. doi:10.1016/j.childyouth.2010.10.001

Ladany, N., Hill, C. E., Corbett, M. M., & Nutt, E. A. (1996). Nature, extent, and importance of what psychotherapy trainees do not disclose to their supervisors. *Journal of Counseling Psychology, 43*, 10-24. doi:10.1037/0022-0167.43.1.10

Ladany, N., Lehrman-Waterman, D., Molinaro, M., & Wolgast, B. (1999). Psychotherapy supervisor ethical practices: Adherence to guidelines, the supervisory working alliance, and supervisee satisfaction. *The Counseling Psychologist, 27*(3), 443-475. doi:10.1177/0011000099273008

Landsman, M. (2001). Commitment in public child welfare. *Social Service Review, September, 2001*, 386-419. doi:10.1086/322857

Lehrman-Waterman, D., & Ladany, N. (2001). Development and validation of the evaluation process within supervision inventory. *Journal of Counseling Psychology, 48*(2), 168-177. doi: 10.1037//0022-0167.48.2.168

Lindsey, D. (2004). *The welfare of children*. New York, NY: Oxford University Press.

McDonald, R. (1999). *Test theory: A unified treatment*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Mertler, C. A., & Vannatta, R. A. (2005). *Advanced and multivariate statistical methods* (3rd ed.). Glendale, CA: Pyrczak.

Mor Barak, M. E., Nissly, J. A., & Levin, A. (2001). Antecedents to retention and turnover among child welfare, social work, and other human service employees: What can we learn from past research? A review and meta-analysis. *Social Services Review, December*, 625-661. doi:10.1086/323166

Mor Barak, M. E., Levin, A., & Nissly, J. (2006). Why do they leave? Modeling child welfare workers' turnover intentions. *Children & Youth Services Review, 28*(5), 548-577. doi:10.1016/j.childyouth.2005.06.003

Munson, C. E. (1976). Professional autonomy and social work supervision. *Journal of Education for Social Work, 12*(Fall), 95-102.

Munson, C. E. (2002). *Handbook of clinical social work supervision* (3rd ed.). Binghamton, NY: Haworth Press.

Murphy, S. Y., & Goodson, J. (2007). Numbers alone do not tell the whole story: A program evaluation designed to generate evidence-based knowledge and practice in child welfare supervision. *Journal of Evidence-Based Social Work, 4*(3/4), 81-101. doi:10.1300/J394v04n03_06

National Association of Social Workers (2004). *If you're right for the job, it's the best job in the world.* Washington, DC: Author.

National Child Welfare Resource Center for Organizational Improvement (2007). CFSR training and technical assistance package. Retrieved from http://muskie.maine.edu/helpkids/

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Oaster, T. R. (1989). Number of alternatives per choice point and stability of Likert-type scales. *Perceptual and Motor Skills, 68*, 549-550. doi:10.2466/PMS.68.2.549-550

Pearlman, L. A., & Saakvitne, K. W. (1995). Treating therapists with vicarious traumatization and secondary traumatic stress disorders. In C. Figley (Ed.), *Compassion fatigue: Coping with secondary traumatic stress disorder in those who treat the traumatized* (pp. 150-177). New York, NY: Brunner/Mazel.

Perry, R. (2004). Factors influencing MSW students' interest in public child welfare. *Journal of Human Behavior in the Social Environment, 10*(2), 1-31. doi: 10.1300/J137v10n02_01

Perry, R. (2006a). Do social workers make better child welfare workers than non-social workers? *Research on Social Work Practice, 16*(4), 392-405. doi:10.1177/1049731505279292

Perry, R. (2006b). Education and child welfare supervisor performance: Does a social work degree matter? *Research on Social Work Practice, 16*(4), 591-604. doi:10.1177/1049731506290548

Poertner, J., & Rapp, C. A. (1983). What is social work supervision? *The Clinical Supervisor, 1*(2), 53-67.

Renner, L. M., Porter, R. L., & Preister, S. (2009). Improving retention of child welfare workers by strengthening skills and increasing supports for supervisors. *Child Welfare, 88*(5), 109-127.

Rhodes, R. E., Matheson, D. H., & Mark, R. (2010). Evaluation of social cognitive scaling response options in the physical activity domain. *Measurement in Physical Education and Exercise Science, 14*, 137-150. doi:10.1080/1091367X.495539

Rubin, A. & Babbie, E. (2001). *Research methods for social work* (4th ed.). Belmont, CA: Wadsworth.

Ryan, J. P., Garnier, P., Zyphur, M., & Zhai, F. (2006). Testing the effects of caseworker characteristics in child welfare. *Children and Youth Services Review, 28,* 993-1006. doi:10.1016/j.childyouth.2005.10.013

Rycraft, J. (1994). The party isn't over: The agency role in the retention of public child welfare caseworkers. *Social Work, 39*(1). 75-80.

Salas, E., & Cannon-Bowers, J. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52,* 471-499. doi:10.1146/annurev.psych.52.1.471

Salus, M. (2004). *Supervising child protective services caseworkers*. Washington, DC: United States Department of Health and Human Services [HHS], Administration for Children and Families [ACF], Administration on Children, Youth and Families, Children's Bureau and Office on Child Abuse and Neglect.

Samantrai, K. (1992). Factors in the decision to leave: Retaining social workers with MSWs in public child welfare. *Social Work, 37,* 454-458.

Scannapieco, M. & Connell-Corrick, K. (2003). Do collaborations with schools of social work make a difference for the field of child welfare? Practice, retention and curriculum. *Journal of Human Behavior in the Social Environment, 7*(1/2), 35-51.

Schorr, A. (2000). The bleak prospect for public child welfare. *Social Services Review, March 2000,* 124-136.

Scott, D., & Farrow, J. (1993). Evaluating standards of social work supervision in child welfare and hospital social work. *Australian Social Work, 45*(2), 33-41.

Shackelford, K. K., & Payne, N. G. (2003). AL-MS child protective service casework supervision. *Professional Development: The International Journal of Continuing Social Work Education, 6(*1/2), 40-46.

Shulman, L. (1993). *Interactional supervision*. Washington, DC: NASW Press.

Shulman, L., Robinson, E., & Luckj, A. (1981). *A study of content, context, and skills of supervision*. Vancouver, Canada: University of British Columbia.

Smith, B. (2005). Job retention in child welfare: Effects of perceived organizational support, supervisor support, and intrinsic job value. *Children and Youth Services Review, 27*(2), 153-169. doi:10.1016/j.childyouth.2004.08.013

Spector, P. E. (1992). *Summated rating scale construction.* [Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-082]. Newbury Park, CA: Sage.

Steppe, S., & Jones, J. L. (2007). Longitudinal research design and the realities of changing practice environments: The difficulty of testing models of evidence based practice—a case study. *Journal of Evidence-Based Social Work, 4*(3/4), 133-143. doi:10.1300/J394v04n03_9

Stewart, N. (2008, July 15). Another death comes as D.C. probes caseworker backlogs. *The Washington Post.* Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2008/07/14/AR2008071401044_pf.html

Strand, V., & Badger, L. (2005). Professionalizing child welfare: An evaluation of a clinical consulting model for supervisors. *Child and Family Services Review, 27,* 865-880. doi:10.1016/j.childyouth.2004.12.001

Strand, V., & Badger, L. (2007). A clinical consultation model for child welfare supervisors. *Child Welfare, 86*(1), 79-97. doi:0009^021 /2007/01079-96

Sundet, P. A., Mermelstein, J., & Watt, J. W. (2003). The role demonstration model of supervision. *Professional Development: The International Journal of Continuing Social Work Education, 6*(2), 60-67.

Tracy, E. M., & Pine, B. (2000). Child welfare education and training: future trends and influences. *Child Welfare, 79*(1), 115-123. Retrieved from http://www.cinahl.com/cgi-bin/refsvc?jid=699&accno=2000015895

Travis, D. J., & Mor Barak, M. E. (2010). Fight or flight? Factors influencing child welfare workers' propensity to seek positive change or disengage from the jobs. *Journal of Social Service Research, 36*(3), 188-205. doi:10.1080/01488371003697905

Tsui, M. S. (1997). Empirical research on social work supervision: The state of the art (1970-1995). *Journal of Social Service Research, 23*(2), 39-54.

Tsui, M. S. (2005). *Social work supervision: Contexts and concepts*. Thousand Oaks, CA: Sage.

Tsui, M. S., & Ho, W. S. (1997). In search of a comprehensive model of social work supervision. *The Clinical Supervisor, 16*(2), 181-205.

United States Department of Health and Human Services, Administration for Children and Families, Children's Bureau (n.d. a). *Child and Family Service Review Fact Sheet*. Retrieved from http://www.acf.hhs.gov/programs/cb/cwmonitoring/recruit/cfsrfactsheet.htm

United States Department of Health and Human Services, Administration for Children and Families, Children's Bureau (n.d. b). *Summary of the result of the 2001-2004 child and family services review*. Retrieved from http://www.acf.hhs.gov/programs/cb/cwmonitoring/results/genfindings04/genfindings04.pdf

United States Department of Health and Human Services, Administration for Children and Families, Children's Bureau (2007). *Guidance to states on actions they are required to take for the June 30, 2007 submission of the Annual Progress and Services Report (APSR) required under Title IV-B of the Social Security Act (the Act), the Child Abuse Prevention and Treatment Act (CAPTA), and the Chafee Foster Care Independence Program (CFCIP), including the ETV program; the submission of the CFS-101, Part I, Annual Budget Request, Part II, Annual Summary of Child and Family Services, and Part III, Annual Expenditure Report-Title IV-B, Subparts 1 and 2, Chafee Foster Care Independence Program (CFCIP) and Education and Training Vouchers (ETV)* [ACYF-CB-PI-07-05]. Retrieved from http://www.acf.hhs.gov/programs/cb/laws_policies/policy/pi/2007/pi0705.htm

United States Department of Health and Human Services, Administration for Children and Families, Children's Bureau (2008). *Guidance to states on actions they are required to take for the June 30, 2008 submission of the Annual Progress and Services Report (APSR) required under Title IV-B of the Social Security Act (the Act), the Child Abuse Prevention and Treatment Act (CAPTA), and the Chafee Foster Care Independence Program (CFCIP), including the ETV program; the submission of the CFS-101, Part I, Annual Budget Request, Part II, Annual Summary of Child and Family Services, and Part III, Annual Expenditure Report—Title IV-B, Subparts 1 and 2, CFCIP and ETV* [ACYF-CB-PI-08-03]. Retrieved from http://www.acf.hhs.gov/programs/cb/laws_policies/policy/pi/2008/pi0803.htm

United States Department of Health and Human Services Administration for Children and Families Children's Bureau (2009). *Fiscal year 2010 allotments for states, territories and insular areas Title IV-B subpart 2: Monthly caseworker visit*. Retrieved from http://www.acf.hhs.gov/programs/cb/laws_policies/policy/pi/2010/pi2009a3.pdf

United States Department of Health and Human Services, Administration for Children and Families, Children's Bureau (2010 a). *The AFCARS Report, Preliminary FY 2009 estimates as of July 2010, #17*. Retrieved from http://www.acf.hhs.gov/programs/cb/stats_research/afcars/tar/report17.htm

United States Department of Health and Human Services, Administration for Children and Families, Children's Bureau (2010 b). *Child and family service review technical bulletin #5*. Retrieved from http://www.acf.hhs.gov/programs/cb/cwmonitoring/general_info/tech_bull5.pdf

United States Department of Health and Human Services, Administration for Children and Families, Children's Bureau (2010 c). *Child maltreatment 2009*. Retrieved from http://www.acf.hhs.gov/programs/cb/pubs/cm09

United States Department of Health and Human Services, Administration for Children and Families, Children's Bureau (2011). *FY 2012 justification of estimates for appropriation committees*. Retrieved from http://www.acf.hhs.gov/programs/cb/laws_policies/policy/pi/2010/pi1009a3.pdf

United States General Accounting Office (2003). *Child welfare: HHS could play a greater role in helping child welfare agencies recruit and retain staff* GAO-03-357. Retrieved from http://www.gao.gov/new.items/d03357.pdf

United States General Accounting Office (2006). *Improving social service program, training, and technical assistance information would address long-standing service-level and workforce challenges* GAO-07-75. Retrieved from http://www.gao.gov/new.items/d0775.pdf

United States General Accounting Office (2007). *Child welfare: Additional federal actions could help states address challenges in providing services to children and families* GAO-07-850T. Retrieved from http://www.gao.gov/new.items/do780t.pdf

Virginia Department of Social Services (2007). *Foster care manual*. Retrieved from http://www.dss.virginia.gov/family/fc/manual.cgi

Virginia Department of Social Services (July, 2008). *Broadcast 5056: Legislative mandate for minimal training of foster care and adoption local staff and supervisors.* Retrieved from http://www.localagency.dss.state.va.us/ broadcasts/docs/2008/07/15/fc_7_15_08.pdf

Virginia Department of Social Services (August, 2008). Broadcast 5057: Monthly caseworker contacts with children in foster care. Retrieved from http://www.localagency.dss.state.va.us/broadcasts/docs/2008/08/15/ fc_8_15_08.pdf

Virginia Department of Social Services (2008). *Child protective services manual.* Retrieved from http://www.dss.virginia.gov/family/cps/manual.cgi

Wagner, R., van Reyk, P., & Spence, N. (2001). Improving the working environment for workers in children's welfare agencies. *Child and Family Social Work, 6,* 161-178.

Watson, K. W. (1973). Differential supervision. *Social Work, 8*(3), 37-43.

Wehrman, K. C., Shin, H., & Poertner, J. (2002). Transfer of training: An evaluation study. In K. Briar-Lawson & J. L. Zlotnik (Eds.), *Evaluation research in child welfare: Improving outcomes through university-public agency partnerships* (pp. 23-38). Philadelphia, PA: Haworth Press.

Woods, M. E., & Hollis, F. (2000). *Casework: A psychosocial therapy* (5th ed.). Boston, MA: McGraw-Hill.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 806-838. doi:10.1177/0011000006288127

Wright, L. (2003). Training in its organizational context: Sharing responsibility. *Performance Improvement Quarterly, 16*(4), 5-22.

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research, 25*(5), 508-518. doi:10.1177/0193945903252998

Zlotnik, J. L. (2002). Preparing social workers for child welfare practice: Lessons from a historical review of the literature. *Journal of Health and Social Policy, 15*, 5-21.

Zlotnik, J. L. (2003). The use of Title IV-E training funds for social work education: An historical perspective. *Journal of Human Behavior in the Social Environment, 7*, 5-20.

Zlotnik, J. L. (2006). No simple answers to a complex question: A response to Perry. *Research on Social Work Practice, 16*(4), 414-416. doi:10.1177/1049731505283701

Zlotnik, J. L., DePanfilis, D., Daining, C., & Lane, M. M. (2005). *Factors influencing retention of child welfare staff: A systemic review of research*. Washington, DC: Institute for the Advancement of Social Work Research.

Appendix A

RESEARCH SUBJECT INFORMATION AND CONSENT FORM

TITLE: Experience of Supervision Scale: The development of an instrument to measure child welfare workers' experience of supervision (Part I)

VCU IRB NO.:

This consent form may contain words that you do not understand. Please ask the study staff to explain any words that you do not clearly understand. You may take home an unsigned copy of this consent form to think about or discuss with family or friends before making your decision.

PURPOSE OF THE STUDY

The purpose of the present study is to increase understanding about the process of child welfare supervision by developing a standardized instrument which measures supervision through analysis of supervisory behaviors from the child welfare worker's perspective. A standardized instrument will allow for comparison between studies and a generalization of findings in order to develop better supervisory training programs, to improve the quality of child welfare supervision, in general, and to improve the quality of services provided to children and family served by the child welfare system.

You are being asked to participate in this study because you are a child welfare worker or child welfare supervisor. You know a lot about supervision in child welfare and your ideas about supervision are important to the process of developing the Experience of Supervision Scale.

DESCRIPTION OF THE STUDY AND YOUR INVOLVEMENT

If you decide to be in this research study, you will be asked to sign this form after you have had all your questions answered and understand what will happen to you.

In this study you will be asked to attend one focus group or participate in an individual interview. Focus groups are expected to last approximately 45 minutes. In each focus group you will be in a group with 7 or 8 other child welfare workers or supervisors. In this group, you will be asked to talk about supervision is general. Focus group participants will be asked to help generate a comprehensive list of supervisory activities. Idea will be written down by the researcher, but your name will not be included in any notes from the group. The group will not be tape recorded and there will be no record beyond the researcher's notes of your specific contributions to the group. If you are interviewed instead of participating in a group, your interview will also not be

recorded. The researcher's notes will not include your name and your identity will not be associated in any way with your responses.

RISKS AND DISCOMFORTS
It is not anticipated that talking about supervision in general will create any psychological or emotional discomfort. However, you do not have to talk about any subjects that you would prefer not to address and you can leave the focus group or stop the interview at any time.

BENEFITS TO YOU AND OTHERS
You may not get any direct benefit from this study, but, the information learned from people in this study may eventually lead to better supervisory training programs and improved supervision in the field of child welfare.

Depending on when and where your focus group is scheduled, you may be provided with refreshments.

COSTS
There are no costs for participating in this study other than the time you will spend in the group.

ALTERNATIVES
There is no alternative to this study other than choosing not to participate. Your participation is completely voluntary and you may choose not to participate at any time.

CONFIDENTIALITY
Information you share during focus groups or in an interview could be identified as your words, but in this study, every effort will be made to keep information shared anonymous. Neither focus groups nor interviews will be recorded. The researcher will take notes detailing your responses, but your name will not be used in the notes. The information generated in the groups will be primarily used to generate items to be included in the Experience of Supervision scale, to be piloted in a second study. The format requirements of a scale mean that all ideas generated in the group will have to be re-written to fit the format and, therefore, are unlikely to be identifiable.

Information about this process and the data generated through the focus groups may be presented at meetings or published in papers, but your name will not ever be used in these presentations or papers. If the researcher chooses to report on the focus group process and the data generated through this process, all data will be presented without identifying information.

We will not tell anyone the answers you give us; however, information from the study and information from the consent form signed by you may be looked at or copied for research or legal purposes by the Virginia Commonwealth University

IF AN INJURY HAPPENS
It is not anticipated that anyone would be hurt by participating in this study. However, if you are injured because of being in this study, tell the researcher right away. Arrangements will be made for short-term emergency care or referral if it is needed.

Bills for treatment may be sent to you or your insurance. Your insurance may or may not pay for taking care of injuries that happen because of being in this study.

VOLUNTARY PARTICIPATION AND WITHDRAWAL
You do not have to participate in this study. If you choose to participate, you may stop at any time without any penalty. You may also choose not to answer particular questions that are asked in the study.

QUESTIONS
In the future, you may have questions about your participation in this study. If you have any questions, complaints, or concerns about the research, contact:

Em Parente, LCSW, PhD candidate
434-970-3579
parenteme@vcu.edu

If you have any questions about your rights as a participant in this study, you may contact:

Office for Research
Virginia Commonwealth University
800 East Leigh Street, Suite 113
P.O. Box 980568
Richmond, VA 23298
Telephone: 804-827-2157

You may also contact this number for general questions, concerns or complaints about the research. Please call this number if you cannot reach the research team or wish to talk to someone else. Additional information about participation in research studies can be found at http://www.research.vcu.edu/irb/volunteers.htm.

CONSENT

I have been given the chance to read this consent form. I understand the information about this study. Questions that I wanted to ask about the study have been answered. My signature says that I am willing to participate in this study. I will receive a copy of the consent form once I have agreed to participate.

Participant name printed	Participant signature	Date

_____

Name of Person Conducting Informed Consent Discussion / Witness (Printed)

_____ _____

Signature of Person Conducting Informed Consent Discussion / Witness — Date

_____ _____

Principal Investigator Signature (if different from above) — Date

Appendix B

Focus Group Script and Discussion Questions

As you know, I am working on developing an instrument to measure supervisory activities or specific supervisory behaviors so that at some point in the future we can better understand how these activities or behaviors impact or influence other things- like worker moral, worker retention, and client outcomes.

I am asking for your help to develop a comprehensive list of things that supervisors do: things they say, the activities they engage in, the things you see them doing and the things you know they do, but happen behind the scenes.

There aren't right or wrong answers. Some ideas may be similar, that's okay. I am more worried about leaving something important out than I am about having too many suggestions.

I would also like to emphasize that I am not asking about any particular supervisor- supervisee relationship. I am asking about supervisors in general. I hope that everyone will feel comfortable talking about their experiences with supervision, in general, but if at any point, you do not feel comfortable sharing, you certainly don't have to. Additionally, I would ask that as much as possible, we all avoid inferring from individual's comments who they might be talking about. Our topic is supervision in general.

Thank you again for being willing to participate.

1. Let's begin by developing a list of all the things you can think of that supervisors do.

2. There is a model of casework supervision that was developed by Alfred Kadushin. Kadusin separates supervisory functions into three groups: administrative, educational and supportive. I'd like to talk about each of these functions and see if it helps us to come up with more ideas

   Administrative supervision is operationally defined as those supervisory practices concerned with "the correct, effective, and appropriate implementation of agency policies and procedures" (Kadushin & Harkness, 2002, p. 20). Administrative supervision includes activities like conducting performance evaluations, disciplinary actions, assigning work, communicating new expectation or policies to workers, etc.

   In thinking about the administrative function of supervision, do additional supervisory activities come to mind?

3. Educational supervision is defined as those supervisory practices intended to improve worker skills and capacities; the primary goal of educational supervision is "to dispel ignorance and upgrade skill" (Kadushin & Harkness, 2002, p. 20). Educational supervision includes such activities as case consultation, providing examples, teaching skills and practicing skills.

   In thinking about the educational function of supervision, do additional supervisory activities come to mind?

4. The primary goal of supportive supervision is "to improve morale and job satisfaction" (Kadushin & Harkness, 2002, p. 20). Supportive supervision is defined as those supervisory practices which help the supervisee feel good about his or her job. Supportive activities include helping the worker to have reasonable expectations for him or herself, helping the worker recognize when personal history or feelings are influencing his or her perceptions of work experiences, and facilitating the recognition of small successes and good work.

   In thinking about the supportive function of supervision, do additional supervisory activities come to mind?

5. Before we conclude, have you thought of anything else that we should include?

Appendix C

Experience of Supervision Scale Survey

Appendix D

Director/Asst. Director Letter (sent via e-mail)

Subject line: Child Welfare supervision research project

Body:

Dear [Director or Assistant Director],

My name is Em Parente and I am a PhD candidate at the School of Social Work at Virginia Commonwealth University. I am also a Foster Care Supervisor with the Charlottesville Department of Social Services.

For my dissertation project, I am attempting to develop a scale to measure child welfare workers' experience of supervision. A standardized instrument will allow for comparison between studies and a generalization of findings in order to develop better supervisory training programs, to improve the quality of child welfare supervision, in general, and to improve the quality of services provided to children and families served by the child welfare system.

I have developed a draft of a scale based on theoretical literature on social work supervision, previous research on supervisory behaviors, and focus groups with child welfare workers and supervisors. Now, I would like to invite the child welfare workers in your agency to complete a survey which will help me refine the Experience of Supervision Scale.

I am hoping that you will provide me with the name of an individual within your agency who can be my point of contact. Participants will access the survey through a web-link provided in an email. I would like to send my email to the contact person so that they can forward it to all of your child welfare workers.

As an incentive for participation, workers will have an opportunity to enter into a drawing for six $50 gift cards.

The VCU Office of Research Subjects Protection has granted IRB approval for this study [study #]. If you have any questions or concerns about this study, please feel free to contact me at parente@charlottesville.org or 434-970-3579.

If you could please e-mail me the contact information for the individual who can be my point of contact at your agency, it would be greatly appreciated. If you do not want me to invite the child welfare workers of your agency to participate, please let me know that as well. I will follow up this email within a week with a phone call to provide you with an opportunity to ask any questions you might have.

Thank you for your time.

Em Parente

Appendix E

Contact Person Letter (sent via e-mail)

Subject line: Research project

Body:

Dear [name of contact person],

Thank you so much for being willing to be my point of contact in your agency.

My name is Em Parente and I am a PhD candidate at the School of Social Work at Virginia Commonwealth University. I am also a Foster Care Supervisor with the Charlottesville Department of Social Services.

For my dissertation project, I am attempting to develop a scale to measure child welfare workers' experience of supervision. I have developed a draft of a scale based on theoretical literature on social work supervision, previous research on supervisory behaviors, and focus groups with child welfare workers and supervisors. Now, I would like to invite the child welfare workers in your agency to complete a survey which will help me refine the Experience of Supervision Scale.

As an incentive for participation, workers will have an opportunity to enter into a drawing for six $50 gift cards.

I will follow this email with an email invitation to participate in the study. The email can be forwarded "as is" to the child welfare workers in your agency.

In a couple of weeks, I will get in touch with you again and send you another email to send out as a reminder to the workers.

Thank you so much for helping me with this. If you have any questions or concerns about this study, please feel free to contact me at parente@charlottesville.org or 434-970-3579.

Thank you again.

Em Parente

Appendix F

Invitation to participate in the study
(sent via email to child welfare workers by the contact person)

Subject line: Will you help me to develop a scale to measure supervision in child welfare?

Body:

Hi. My name is Em Parente and I am a PhD candidate at the School of Social Work at VCU. I am also a Foster Care Supervisor with the Charlottesville Department of Social Services.

For my dissertation project, I am attempting to develop a scale to measure child welfare workers' experience of supervision. Currently, I have a draft Experience of Supervision scale, but I need input from child welfare workers to refine the draft and to evaluate how well the scale works.

I've set up a survey in Survey Monkey to collect your input. I hope you will choose to participate. The survey should take approximately 15 minutes to complete. At the end of the survey you will have an opportunity to provide your email address to be entered into a drawing for a $50 Best Buy gift card. I will draw for 3 cards at the end of April and 3 more at the end of May. The sooner you complete the survey the better your chance to win.

Here is the link to the survey: http://www.surveymonkey.com/s/2D6PJCP

If you have any questions about this project, please feel free to contact me at parente@charlottesville.org or (434) 970-3579.

If you have any questions about your rights as a participant in this study, you may contact:

> Office for Research Subjects Protection
> Virginia Commonwealth University
> 800 East Leigh St., Suite 114
> PO Box 980568
> Richmond, VA 23298
> 804-828-0868

Thank you,
Em

Appendix G

Reminder invitation to participate in the study
(sent via email to child welfare workers by the contact person)

Subject: A reminder about the child welfare supervision scale research project

Body:

Hi. A couple weeks ago you received an email asking for your help to develop a scale to measure child welfare supervision. If you've already responded, terrific. If you would still like to respond, there is still time. The survey should only take about 15 minutes to complete.

Here is the link to the survey: http://www.surveymonkey.com/s/2D6PJCP

Just as a reminder, everybody who takes the survey will have an opportunity to be entered in to a drawing for a $50 Best Buy gift card. I will draw for 3 winners at the end of April and 3 more at the end of May. The sooner you complete the survey the better your chance to win.

If you have any questions about this project, please feel free to contact me at parente@charlottesville.org or (434) 970-3579.

If you have any questions about your rights as a participant in this study, you may contact:

Office for Research Subjects Protection
Virginia Commonwealth University
800 East Leigh St., Suite 114
PO Box 980568
Richmond, VA 23298
804-828-0868

Thank you,

Em

Appendix H

First Draft of Item Stems

1. Meets regularly for supervision
2. Communicates that she values me
3. Takes workloads into account when assigning cases
4. Double-checks my paperwork for accuracy
5. Critiques my documentation to improve quality and completeness
6. Routinely provides feedback
7. Encourages professional growth
8. Suggests trainings I might attend
9. Teaches me new skills
10. Uses role play to help me practice new skills
11. Identifies a senior worker or mentor to provide additional support to me as needed
12. Models professionalism with clients
13. Models good practice skills with clients
14. Goes into the field with me to provide support with challenging situations
15. Uses observations of my work in the field to help me improve my practice skills
16. Pitches in and helps handle case emergencies
17. Helps me understand what about particular cases are personally challenging
18. Mediates conflicts in case situations
19. Addresses conflicts between program areas in the workplace
20. Advocates for better salaries for staff
21. Advocates for systems intervention to maintain appropriate case loads for staff
22. Recognizes individual accomplishments publically
23. Recognizes group or team accomplishments publically
24. Tells me when I've done a good job
25. Provides encouragement when I'm having a hard time
26. Establishes deadlines
27. Monitors progress towards deadlines
28. Helps interpret policy as it applies to my cases
29. Consults with specialists or attorney when clarification regarding policy is needed
30. Encourages me to use policy to find answers for myself
31. Provides guidance regarding crucial case decisions
32. Shares responsibility when crucial case decisions are challenged
33. Takes an interest in me as a person
34. Facilitates good team work
35. Makes expectations clear

36. Helps me to set performance goals for myself
37. Actively involves me in the performance evaluation process
38. Encourages me to continue to develop as a worker
39. Helps me to think clinically about my cases
40. Is available for consultation when I have a case crisis
41. Helps me understand the larger agency and state context within which I work
42. Is open to feedback about her own performance
43. Asks if supervision is meeting my needs
44. Treats me fairly
45. Advocates for policies which help keep me safe
46. Arranges coverage so I can go to training
47. Encourages me to take care of myself
48. Challenges me when I have unreasonable expectations for clients
49. Challenges me when I have unreasonable expectations about my own performance
50. Is supportive when I need support
51. Provides me with direction when I need direction
52. Is trustworthy
53. Encourages me to develop areas of expertise
54. When assigning cases is sensitive to the kinds of cases I prefer
55. Models problem-solving skills
56. Acknowledges my skills and abilities
57. Models using humor appropriately to maintain perspective
58. Acknowledges when she makes a mistake
59. Values my contributions to problem solving
60. Keeps me informed about trends and issues which affect my work
61. Advocates for resource development to address resource gaps
62. Filters all the policy and practice changes so I get exactly what I need
63. Strengthens collaborative relationships with community partner agencies
64. Provides a safe place to talk about feeling overwhelmed
65. Acknowledges when I'm having a bad day
66. Models appropriate personal/professional boundaries
67. Answers my questions regarding policy expertly
68. Listens to me explain the situation/issue before providing input
69. Pitches in when I need help
70. Uses own experience doing the job to show me how to do the job better
71. Covers court hearing for me when I am unavailable
72. Comes to court to support me when I have a challenging court situation
73. Lets me know when there is something I need to improve
74. Keeps up with all the changes so she knows how to do the job I am doing
75. Encourages me to manage my workload in the manner that works best for me
76. Backs me up if a decision I've made gets challenged
77. Is honest
78. Has a strong collaborative relationship with other supervisors
79. Works as hard as she expects me to
80. Manages administrative workload so that she is still available to me

81. Knows me (my strengths, my skills and my needs)
82. Trusts me to do a good job
83. Understands why I am not always able to get everything done on time
84. Encourages me to prioritize my tasks
85. Makes me feel like how I'm doing is important to her
86. Gives positive feedback
87. Recognizes when I've done something well
88. Helps me do a better job when I am struggling rather than taking over the case
89. Arranges coverage so that I can take time off
90. Does not share the private things I say in conference with others
91. Arranges for others to pitch in when I need help
92. Helps me to recognize when a particular case is really stressing me out
93. Follows up with me regarding cases I'm worried about
94. Hires new workers who fit well in my team
95. Supports flexible scheduling to meet client needs and personal needs
96. Models commitment to ongoing professional development by engaging actively in opportunities to learn
97. Is professionally educated
98. Promotes my involvement in workgroups and special projects as professional development opportunities for me
99. Encourages more seasoned workers to attend more specialized trainings
100. Encourages me to develop a practice specialization
101. Arranges for workers to share what they've learned at training with the rest of the team
102. Makes opportunities for me to have case consultation with my team
103. Participates with me in family meetings for challenging cases
104. Takes action when I have a case emergency
105. Provides up to date info on local resources and providers
106. Gives me time to absorb policy or practice changes before expecting me to act on them
107. Gives me time to work through my reaction to bad news before expecting me to be constructive in dealing with it
108. Applies policy consistently across cases and case situations
109. Remembers cases from conference to conference so doesn't have to be told all the details again
110. Manages her personal life so it doesn't interfere with her availability
111. Reminds me of deadlines
112. Holds me accountable to complete my work on time
113. Models taking responsibility for her own work
114. Is willing to intervene directly with clients who are angry with me if I need her help
115. Advocates for reasonable caseloads and workloads with management
116. Implements strategies or develops resources to help manage unreasonable workloads
117. Acknowledges when my caseload/workload is particularly heavy
118. Addresses individual performance issues which are negatively impacting the team
119. Has fair performance expectations across the team
120. Advocates effectively with management on behalf of the team

121. Use case consultation to encourage me to connect training I have received to specific case situations and/or my intervention efforts
122. Provides examples when teaching a skill
123. Balances encouraging self awareness and respecting my privacy
124. Supports taking time off to deal with family emergencies
125. Encourages me to prepare for career advancement
126. Notices when I have gotten better at doing something
127. Expects me to be successful
128. Pushes me so that I can excel
129. Communicates clearly about her expectations
130. Encourages me to express my own ideas about cases
131. Balances my need for specific suggestions with my need to figure things out for myself
132. Addresses misconceptions about child welfare in the community
133. Encourages me to think for myself
134. Helps me to explore where my ideas about a case come from
135. Appropriately challenges any personal bias which affects my work
136. Expects me to set performance goals
137. Acknowledges when I have made progress towards performance goals
138. Monitors my documentation and addresses problems as they arise
139. Helps the agency to respond proactively to client trends
140. Works strategically to improve efficiency within the agency
141. Manages new state mandates to minimize the negative impact they will have on me
142. Lets me know it's okay to talk about my mistakes
143. Asks me about what motivates me
144. Lets me know it's okay to take time off to deal with personal/family concerns
145. Encourages me to schedule vacation
146. Stays late when there is an emergency to be on hand
147. Addresses performance issues on an ongoing basis, not just in written evaluations
148. Helps staff get along with each other
149. Matches what she provides in supervision to my level of experience
150. Has a vision for how we are going to keep improving the outcomes for the children and families we serve
151. Is hopeful even when times are hard
152. Is strength oriented regarding clients
153. Is respectful and strengths oriented regarding staff
154. Manages her own stress well
155. Helps me to see how what I'm doing is making a difference
156. Allows me to make choices about how I do the work
157. Encourages me to balance work and personal life appropriately
158. Empowers me to make my own case decisions
159. Speaks to me privately when I've made a mistake
160. Makes an effort to understand my side of the story when there is a complaint
161. Does what she says she is going to do
162. Says what she means
163. Models ethical behavior with clients and staff

164. Provides opportunities for trying new things
165. Provides opportunity to observe other areas of practice
166. Is committed to the field of child welfare
167. Likes the clients
168. Is enthusiastic about the work
169. Values the development of clinical skills
170. Advocates for adequate physical resources to do the job (computer, car, cell phone, etc)
171. Advocates for better benefits and pay for workers
172. Avoid being overly focused on minor details

Appendix I

Initial Items (sorted) for use with Expert Panel

Administrative

1. Helps the agency to respond proactively to client trends
2. Works strategically to improve efficiency within the agency
3. Manages new state mandates to minimize the negative impact they will have on me
4. Addresses performance issues on an ongoing basis, not just in written evaluations
5. Acknowledges when I have made progress towards performance goals
6. Advocates for adequate physical resources to do the job (computer, car, cell phone, etc)
7. Avoids being overly focused on minor procedural details
8. Double-checks my paperwork for accuracy
9. Critiques my documentation to improve quality and completeness
10. Addresses conflicts between program areas in the Department
11. Advocates for better salaries for staff
12. Advocates for systems intervention to maintain reasonable case loads for staff
13. Monitors progress towards deadlines
14. Helps interpret policy as it applies to my cases
15. Consults with specialists or attorney when clarification regarding policy is needed
16. Provides guidance regarding critical case decisions
17. Facilitates good team work
18. Makes expectations clear
19. Actively involves me in the performance evaluation process
20. Advocates for policies which help keep me safe
21. When assigning cases, is sensitive to the kinds of cases I prefer
22. Keeps me informed about trends and issues which affect my work
23. Advocates for resource development to address resource gaps
24. Filters policy and practice changes so I get exactly the information I need to do my job
25. Strengthens collaborative relationships with community partner agencies
26. Lets me know when there is something I need to improve
27. Has a strong collaborative relationship with other supervisors
28. Hires new workers who fit well in my team
29. Provides up to date info on local resources and providers
30. Gives me time to absorb policy or practice changes before expecting me to act on them
31. Applies policy consistently across cases and case situations
32. Holds me accountable for completing my work on time
33. Implements strategies or develops resources to help manage unreasonable caseloads

34. Addresses individual performance issues which are negatively impacting the team
35. Has fair performance expectations across the team
36. Advocates effectively with management on behalf of the team
37. Makes sure that I am doing my job the way I am supposed to

Supportive

38. Is available for consultation when I have a case crisis
39. Acknowledges when my caseload/workload is particularly heavy
40. Asks if supervision is meeting my needs
41. Is open to feedback about her own performance
42. Shares responsibility when case decisions are challenged
43. Takes an interest in me as a person
44. Tells me when I've done a good job
45. Provides encouragement when I'm having a hard time
46. Helps me understand why a particular case is personally challenging
47. Pitches in and helps handle emergencies
48. Goes into the field with me to provide support with challenging situations
49. Identifies a senior worker or mentor to provide additional support to me as needed
50. Encourages me to take care of myself
51. Challenges me when I have unreasonable expectations about my own performance
52. Acknowledges when she makes a mistake
53. Provides a safe place to talk about feeling overwhelmed
54. Listens to me explain the situation/issue before providing input
55. Covers court hearings for me when I am unavailable
56. Comes to court to support me when I have a challenging court case
57. Keeps up with all the changes so she knows how to do the job I am doing
58. Encourages me to manage my workload in the manner that works best for me
59. Works as hard as she expects me to
60. Is available to me when I have a problem
61. Is accepting when I am not able to get everything done on time
62. Encourages me to prioritize my tasks
63. Arranges coverage so that I can take time off
64. Arranges for others to pitch in when I need help
65. Helps me to recognize when a particular case is really stressing me out
66. Balances encouraging self awareness and respecting my privacy
67. Supports taking time off to deal with family emergencies
68. Asks me about what motivates me
69. Encourages me to take vacation
70. Stays late when there is an emergency to be on hand
71. Lets me know it's okay to talk about my mistakes
72. Helps me to see how what I'm doing is making a difference
73. Encourages me to balance work and personal life
74. Takes time to understand my side of the situation when there is a complaint

75. Remembers cases from conference to conference so she doesn't have to be told all the details again

Educational

76. Models appropriate personal/professional boundaries
77. Empowers me to make case decisions
78. Encourages me to connect training I have received to specific case situations and/or my intervention efforts
79. Provides examples when teaching a skill
80. Models commitment to ongoing professional development by engaging actively in opportunities to learn
81. Promotes my involvement in workgroups and special projects as professional development opportunities for me
82. Arranges for workers to share what they've learned at training with the rest of the team
83. Makes opportunities for me to have case consultation with my team
84. Helps me do a better job when I am struggling rather than taking over the case
85. Uses her own experience doing the job to teach me how to do the job better
86. Encourages me to develop areas of expertise
87. Challenges me when I have unreasonable expectations for clients
88. Arranges coverage so I can go to training
89. Models problem-solving skills
90. Models using humor appropriately to maintain perspective
91. Encourages me to use policy to find answers for myself
92. Uses observations of my work in the field to help me improve my practice skills
93. Encourages me to prepare for career advancement
94. Notices when I have gotten better at doing something
95. Balances my need for specific suggestions with my need to figure things out for myself
96. Helps me to explore where my ideas about a case come from
97. Appropriately challenges any personal bias which affects my work
98. Matches what she provides in supervision to my level of experience
99. Models ethical behavior with clients and staff
100. Provides opportunities to try new things
101. Provides opportunities to observe other areas of practice
102. Suggests trainings I might attend
103. Uses role play to help me practice new skills
104. Models professionalism with clients
105. Models good practice skills with clients
106. Encourages me to develop my strengths
107. Makes opportunities for me to have case consultation with my team
108. Encourages me to have my own ideas about cases

Appendix J

Items with 75% agreement by Expert Panel (Retained)

Administrative

1. Works strategically to improve efficiency within the agency
2. Manages new state mandates to minimize the negative impact they will have on me
3. Addresses conflicts between program areas in the Department
4. Advocates for systems intervention to maintain reasonable case loads for staff
5. Monitors progress towards deadlines
6. Consults with specialists or attorney when clarification regarding policy is needed
7. Advocates for policies which help keep me safe
8. Advocates for resource development to address resource gaps
9. Filters policy and practice changes so I get exactly the information I need to do my job
10. Strengthens collaborative relationships with community partner agencies
11. Has a strong collaborative relationship with other supervisors
12. Applies policy consistently across cases and case situations
13. Holds me accountable for completing my work on time
14. Implements strategies or develops resources to help manage unreasonable caseloads

Supportive

1. Is available for consultation when I have a case crisis
2. Acknowledges when my caseload/workload is particularly heavy
3. Asks if supervision is meeting my needs
4. Takes an interest in me as a person
5. Tells me when I've done a good job
6. Provides encouragement when I'm having a hard time
7. Pitches in and helps handle emergencies
8. Encourages me to take care of myself
9. Provides a safe place to talk about feeling overwhelmed
10. Listens to me explain the situation/issue before providing input
11. Comes to court to support me when I have a challenging court case
12. Is available to me when I have a problem
13. Is accepting when I am not able to get everything done on time
14. Helps me to recognize when a particular case is really stressing me out
15. Balances encouraging self awareness and respecting my privacy

16. Supports taking time off to deal with family emergencies
17. Asks me about what motivates me
18. Encourages me to take vacation
19. Lets me know it's okay to talk about my mistakes
20. Helps me to see how what I'm doing is making a difference
21. Encourages me to balance work and personal life
22. Takes time to understand my side of the situation when there is a complaint
23. *(A) Acknowledges when I have made progress towards performance goals
24. *(A) Facilitates good team work
25. *(A) When assigning cases, is sensitive to the kinds of cases I prefer
26. *(E) Models using humor appropriately to maintain perspective
27. *(E) Notices when I have gotten better at doing something
28. *(E) Encourages me to develop my strengths

<u>Educational</u>

1. Models appropriate personal/professional boundaries
2. Encourages me to connect training I have received to specific case situations and/or my intervention efforts
3. Provides examples when teaching a skill
4. Models commitment to ongoing professional development by engaging actively in opportunities to learn
5. Arranges for workers to share what they've learned at training with the rest of the team
6. Uses her own experience doing the job to teach me how to do the job better
7. Challenges me when I have unreasonable expectations for clients
8. Models problem-solving skills
9. Encourages me to use policy to find answers for myself
10. Uses observations of my work in the field to help me improve my practice skills
11. Helps me to explore where my ideas about a case come from
12. Appropriately challenges any personal bias which affects my work
13. Provides opportunities to try new things
14. Provides opportunities to observe other areas of practice
15. Suggests trainings I might attend
16. Uses role play to help me practice new skills
17. Models professionalism with clients
18. Models good practice skills with clients
19. *(A) Critiques my documentation to improve quality and completeness

*( ) denotes items which were initially sorted by the researcher into another category (A=Administrative; S=Supportive and E=Educational)

Appendix K

Appendix K

Rotated Component Matrix[a]

	Item	Components			Communality
		1	2	3	
Q33	When assigning cases, is sensitive to the kinds of cases I prefer	**.702**	.246	.094	.562
Q46	Asks me about what motivates me	.**684**	.272	.225	.592
Q52	Helps me to recognize when a particular case is really stressing me out	**.670**	.434	.293	.723
Q39	Advocates for resource development to address resource gaps	**.652**	.118	.487	.676
Q57	Provides opportunities to observe other areas of practice	**.649**	.320	.140	.543
Q60	Implements strategies or develops resources to help manage unreasonable caseloads	.**636**	.274	.385	.628
Q47	Encourages me to take vacation	.**635**	.337	.039	.518
Q24	Advocates for systems intervention to maintain reasonable caseloads for staff	**.622**	.235	.415	.614
Q11	Acknowledges when I have made progress towards performance goals	**.610**	.476	.293	.685
Q54	Encourages me to develop my strengths	.**606**	.491	.333	.719
Q36	Uses observations of my work in the field to help me improve my practice skills	.**602**	.267	.392	.685
Q37	Notices when I have gotten better at doing something	**.594**	.492	.309	.691
Q36	Tells me when I've done a good job	.**588**	**.544**	.176	.588
Q16	Uses role play to help me practice new skills	.**584**	.034	.247	.403
Q56	Provides opportunities for me to try new things	**.579**	.389	.239	.544
Q59	Helps me to see how what I'm doing is making a difference	**.557**	.**547**	.275	.684
Q43	Helps me to explore where my ideas about a case come from	.**551**	.286	.437	.577
Q7	Encourages me to connect training I have received to specific case situations and/or my intervention efforts	.**544**	.314	.329	.503

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. [a] Rotation converged in 16 iterations. Factor loadings > .50 are in boldface. Continued next page.

Appendix K (continued)

Rotated Component Matrix[a]

	Item	Components			Communality
		1	2	3	
Q61	Suggests trainings I might attend	**.540**	.253	.340	.471
Q38	Comes to court to support me when I have a challenging court case	**.536**	.140	.215	.353
Q5	Asks if supervision is meeting my needs	**.530**	.335	.351	.517
Q6	Manages new state mandates to minimize the negative impact they will have on me	**.518**	.398	.377	.569
Q58	Lets me know it's okay to talk about mistakes	**.510**	.455	.353	.591
Q8	Provides examples when teaching a skill	**.510**	.450	.436	.653
Q44	Balances encouraging self awareness and respecting my privacy	.498	.425	.376	.571
Q13	Addresses conflicts between program areas in the Department	.491	.446	.468	.659
Q9	Models commitment to ongoing professional development by engaging actively in opportunities to learn	.478	.383	.372	.513
Q10	Arranges for workers to share what they've learned at training with the rest of the team	.420	.237	.358	.361
Q34	Takes the time to understand my side of the situation when there is a complaint	.217	**.730**	.230	.633
Q2	Is available for consultation when I have a case crisis	.123	**.683**	.370	.619
Q29	Provides a safe place to talk about feeling overwhelmed	.466	**.664**	.206	.700
Q14	Takes an interest in me as a person	.424	**.640**	.277	.666
Q30	Facilitates good team work	.400	**.639**	.376	.709
Q20	Pitches in and helps handle emergencies	.307	**.637**	.341	.616
Q42	Is available to me when I have a problem	.221	**.625**	.424	.619
Q19	Provides encouragement when I'm having a hard time	.535	**.615**	.239	.722
Q3	Models appropriate personal/ professional boundaries	.180	**.612**	.450	.610
Q26	Encourages me to take care of myself	.480	**.574**	.273	.635
Q4	Acknowledges when my caseload/workload is particularly heavy	.464	**.566**	.255	.601
Q21	Uses her own experience doing the job to teach me how to do the job better	.429	**.557**	.356	.621
Q50	Is accepting when I am not able to get everything done on time	.353	**.554**	.031	.432

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. [a] Rotation converged in 16 iterations. Factor loadings > .50 are in boldface. Continued next page.

Appendix K (continued)

Rotated Component Matrix[a]

Item		Components			Communality
		1	2	3	
Q15	Encourages me to balance work and personal life	**.539**	**.539**	.266	.652
Q23	Models problem-solving skills	.422	**.537**	.483	.700
Q45	Supports taking time off to deal with family emergencies	.372	**.529**	-.028	.419
Q31	Listens to me explain the situation/issue before providing input	.196	**.523**	.385	.460
Q17	Models good practice skills with clients	.282	.482	.467	.529
Q25	Models using humor appropriately to maintain perspective	.388	.480	.410	.549
Q27	Monitors progress towards deadlines	.272	.164	**.747**	.658
Q55	Holds me accountable for completing my work on time	.236	-.014	**.708**	.557
Q28	Consults with specialists or attorney when clarification regarding policy is needed	.118	.341	**.673**	.583
Q41	Filters policy and practice changes so I get exactly the information I need to do my job	.287	.240	**.643**	.553
Q12	Critiques my documentation to improve quality and completeness	.246	.161	**.630**	.484
Q40	Models professionalism with clients	.106	.529	**.581**	.628
Q49	Has a strong collaborative relationship with other supervisors	.074	.419	**.560**	.495
Q53	Applies policy consistently across cases and case situations	.259	.423	**.556**	.555
Q48	Strengthens collaborative relationships with community partner agencies	.342	.329	**.554**	.532
Q1	Works strategically to improve efficiency within the agency	.368	.378	**.554**	.586
Q32	Advocates for policies which help keep me safe	.361	.304	**.509**	.483
Q51	Appropriately challenges any personal bias which affects my work	.440	.331	.468	.522
Q22	Challenges me when I have unreasonable expectations for clients	.462	.315	.462	.526
Q35	Encourages me to use policy to find answers for myself	.392	.149	.457	.385

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. [a] Rotation converged in 16 iterations. Factor loadings > .50 are in boldface.

Vita

Mary Elizabeth “Em” Parente was born June 20, 1967, in Van Nuys, California, and is a United States citizen. She graduated from Indio High School, Indio, California, in 1985. She earned her Bachelor of Arts in Spanish Literature from Cornell University, Ithaca, New York, in 1989. She then obtained a Master of Social Work degree with a clinical concentration at Boston College, Chestnut Hill, Massachusetts, in 1993. For the last 15 years, she has been working at Charlottesville Department of Social Services, a public child welfare agency, in a variety of positions including Foster Care Worker, Independent Living Coordinator, Student Unit Supervisor and, for the last three years, Foster Care Supervisor.